WARWICK

ECONOMICS

Can Workfare Programs Moderate Conflict?

Evidence from India

Thiemo Fetzer

(This paper also appears as CAGE Discussion paper 436)

August 2019 No: 1220

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Can Workfare Programs Moderate Conflict? Evidence from India *

Thiemo Fetzer

August, 2019

Abstract

Can public interventions persistently reduce conflict? Adverse weather shocks, through their impact on incomes, have been identified as robust drivers of conflict in many contexts. An effective social insurance system moderates the impact of adverse shocks on household incomes, and hence, could attenuate the link between these shocks and conflict. This paper shows that a public employment program in India, by providing an alternative source of income through a guarantee of 100 days of employment at minimum wages, effectively provides insurance. This has an indirect pacifying effect. By weakening the link between productivity shocks and incomes, the program uncouples productivity shocks from conflict, leading persistently lower conflict levels.

Keywords: social insurance, civil conflict, India, NREGA, insurgency

JEL Codes: D74, H56, J65, Q34

> NREGA is the only way forward to take on the Maoists. This is nothing about winning hearts and minds. Its only about giving people work before the rebels come in and convince them that they are a better option than the state. *- NREGA officer, West Midnapore, West Bengal. (BBC 2010)*

*University of Warwick, Department of Economics, Coventry CV4 7AL, United Kingdom. The author is also affiliated with CESifo, CEPR and the The Pearson Institute for the Study and Resolution of Global Conflicts. Support from the Kondrad-Adenauer Foundation, a Bagri Fellowship and an LSE-TISS-TATA grant is gratefully acknowledged. I would like to thank four anonymous referees and the editor for helpful comments and suggestions. I am indebted to Gerard Padro-i-Miquel for his continued support. I would like to thank George Akerlof, Tim Besley, Konrad Buchardi, Jon Colmer, Mirko Draca, Jon de Quidt, Claudia Eger, Maitreesh Ghatak, Solomon Hsiang, Clement Imbert, Sam Marden, Ted Miguel, Hannes Mueller, Torsten Persson, Daniel Sturm, Oliver Vanden Eynde, Juan Vargas, Joachim Voth, Kelly Zhang and seminar participants at LSE, CEP, Oxford Uni, Chicago Harris, Berkeley Goldman School, UPF, CEMFI and NEUDC (2013, 2014) for comments. A special thanks goes to Clement Imbert who shared some of his data.

1 Introduction

The World Bank estimates that 2 billion people currently live in countries where development outcomes are affected by fragility, conflict, and violence. By 2030, the share of global poor living in fragile and conflict-affected areas is projected to reach nearly 50% (Commission on Fragility and Development, 2018). Military and humanitarian interventions costing billions of dollars are aimed at containing the spread of conflicts or alleviating its consequences, but they too often fail to tackle the underlying roots of conflict. Could these resources instead be effectively devoted to public interventions to prevent conflicts from occurring in the first place? If so, how and where should public resources be directed? This study shows that well-designed social insurance policies can play a significant role in fostering stability and building resilience, and, through these channels, help to reduce conflict and fragility.

In order to identify public interventions that can reduce conflict, an understanding of the specific drivers of conflict is needed. Two interlinking empirical regularities stand out in the literature across contexts. The first is the observation that low incomes provide a breeding ground for civil conflict (Collier and Hoeffler, 1998, 2004; Hegre and Sambanis, 2006); the second is the finding that adverse shocks to incomes cause new conflicts to break out, and/or lead existing conflicts to intensify (Bazzi and Blattman, 2014; Dube and Vargas, 2013; Besley and Persson, 2008; Miguel et al., 2004; Fearon and Laitin, 2003). The relationship between income shocks and conflict provides a blueprint for policy. Any public intervention that helps households smooth incomes has the potential to reduce conflict by breaking this key link. This paper provides evidence that social insurance in the form of a public works program in India provided effective insurance against adverse income shocks, thus leading to increased economic resilience and markedly less conflict.

India is an ideal context to study for several reasons. On one side, India suffers from several low-intensity intra-state conflicts in which armed groups fight the state. Existing evidence from India suggests that weather shocks that lead to income losses for people who are uninsured are an important driver of conflict (see Vanden Eynde, 2016; Gawande et al., 2017). On the other hand, India has sufficient state capacity to administer social policies on the ground. In 2006, India introduced the National Rural Employment Guarantee Act (henceforth, NREGA), which established the world's largest public employment program. NREGA provides a safety net by creating a legal entitlement to 100 days of public work per year. Participation is driven by the opportunity cost of employment: the provided work pays the official minimum wage.

As a result, participation is only attractive for individuals with poor outside options, specifically at times when economic conditions are bad. NREGA is the biggest public employment program in history. In the financial year 2011-2012, it reached 3.8% of the world's population, providing employment for 49.8 million rural households, generating 2.114 billion person-days of employment.[1]

The paper presents three main findings: first, I provide evidence that an income channel drives conflict, a finding that confirms results from the existing literature on conflict in India (see Vanden Eynde, 2016; Gawande et al., 2017). Using data on proxies of agricultural income, specifically, agricultural output and agricultural wages, together with a two-period household level panel, I show that prior to the introduction of NREGA, adverse monsoon rainfall realizations are linked to lower agricultural wages, lower agricultural output, and lower household incomes among agricultural laborers. The adverse monsoon rainfall-induced drops in agricultural incomes are, in turn, linked to conflict.

With NREGA being rolled out, these relationships have changed markedly. Specifically, I show that the relationship between monsoon rainfall and agricultural wages has become substantially weaker. Similarly, household incomes among unskilled agricultural laborers effectively cease to be a function of monsoon rainfall. As a result of NREGA, local monsoon variation ceases to affect low skilled agricultural wages, and poorer households are insulated from monsoon-induced income volatility. At the same time, however, monsoon rainfall continues to be an important driver of overall agricultural output (and wages of higher-skilled workers). This is not surprising: while NREGA may provide a stable wage floor, it is not expected to have dramatic effects on the role that monsoon rainfall plays in driving agricultural production processes; similarly, NREGA will not directly affect the wages of higher-skilled laborers, for whom the wage floor established by the program was not binding to begin with. Lastly, I study how the link between monsoon rainfall and conflict changes. The main finding is that the relationship between monsoon rainfall and conflict effectively disappeared. The results are robust to a wide range of checks. Moreover, I am able to provide evidence in support of the underlying common trends assumption: the elasticity linking monsoon rainfall to conflict only becomes distinctly weaker with the roll out of NREGA in a district. I further rule out alternative potential explanations and show that the effect of NREGA is most pronounced in the least-developed districts.

The remaining analysis provides further evidence in support of the reduced-form

[1] See http://nrega.nic.in/netnrega/mpr_ht/nregampr.aspx, accessed on 14.06.2014.

findings, which highlight that the observed changes in conflict stem from the program providing what effectively serves as insurance against adverse monsoon rainfall-induced income shocks. First, using official NREGA data, I show that participation in the program responds to adverse monsoon rainfall realizations, both along the intensive and extensive margin. A quantification exercise suggests that NREGA expenditures flowing into a district compensate for at least 20% of the overall district-level income losses attributable to adverse monsoon rainfall. This estimate is likely *a lower bound* of the actual insurance value of the program to households because the stabilization of agricultural wages offers a further indirect insurance benefit also to individuals not participating in NREGA.[2] Second, I construct a measure capturing whether (or the extent to which) participation in NREGA is responsive to monsoon variation at the district level. I show that districts in which NREGA participation is responsive to monsoon variation experience up to 50% lower levels of conflict. As before, I provide evidence to support for the underlying common trends assumption.

My findings contribute to several literatures. Whether public, non-military interventions affect conflict is an ongoing academic debate. Typically, this literature explores the effect of various shocks to aid. Berman et al. (2011) show that, in particular, small-scale, local development projects can reduce conflict. Nunn and Qian (2014) show that U.S. food aid prolonged low-intensity internal conflict. Their work relates to that of Crost and Johnston (2014) who find a positive relationship between World Bank-funded aid and conflict in the Philippines. Hence, there is mixed evidence on the relationship between aid and conflict. Berman et al. (2013) suggest that key factors are the context in which aid is provided and how it is delivered. NREGA provides a distinctly different delivery vehicle for aid because it is demand-led with self-selection of households ensuring that transfers are targeted. Furthermore, the mechanics of participation makes organized aid-capture of NREGA spoils (as e.g. highlighted in Nunn and Qian, 2014) more difficult; lastly, the decentralized administration ensures that local preferences are reflected in the choice of the public works projects.

This paper also relates to the broader literature on the economics of conflict and labor markets. A core theoretical foundation in this literature is the opportunity cost channel (see Becker, 1968): by reducing the returns to labor, negative productivity shocks may render joining or supporting an insurgency movement incentive compatible. This eventually translates into increased conflict (see Chassang, 2009; Dal Bó

[2] Muralidharan et al. (2017) highlight the general equilibrium effects of NREGA on rural labor markets, suggesting these are likely significantly larger compared to the direct program benefits.

and Dal Bó, 2011). Empirically, evidence in support of this channel is mixed. Bazzi and Blattman (2014) find that positive (commodity price-induced) income shocks may shorten existing conflicts, while Crost and Felter (2019) find that (export) commodity price shocks in the Philippines cause an increase in violence which is concentrated in parts of the country where export crop production is due to large firms. Fetzer and Kyburz (2018) highlight the role of institutions: cohesive democratic institutions in Nigeria weaken the pass-through of commodity price shocks on conflict. Dube and Vargas (2013), while finding evidence consistent with a rapacity-effect also find evidence consistent with an opportunity cost channel as negative shocks to the returns of labor are associated with an intensification of conflict. Shapiro et al. (2011) study how unemployment affects insurgency violence in Afghanistan, Iraq and the Philippines. They find limited support for an opportunity-cost channel. This contrasts with the work of Guardado and Pennings (2017), who find significant evidence in support of its relevance in seasonal labor markets. Iyengar et al. (2011) find that increased public construction spending reduces certain types of labor-intensive insurgent activities. Using a randomized controlled trial, Annan and Blattman (2016) show that providing training and capital to young men, can greatly increase the opportunity cost of becoming a mercenary, and, thus, may contribute to a weakening in the relationship between shocks and conflict.[3] The results presented here suggest that an opportunity cost channel may be driving conflict in India, which becomes weaker once NREGA reduces the pass-through of productivity shocks on to household incomes, particularly among the population most vulnerable to insurgent recruitment and mobilization.

Lastly, the paper relates to a growing literature studying conflict and crime in India. Iyer and Topalova (2014) and Sekhri and Storeygard (2014) study the impact of weather shocks on crimes in general and crimes against minorities specifically, while Vanden Eynde (2016) and Gawande et al. (2017) find evidence in support of an opportunity cost channel driving Maoist conflict. Dasgupta et al. (2017), which is closely related to this paper, use a difference-in-differences approach to compare districts that received NREGA early with those that received it later; they find that the introduction is correlated with a reduction in conflict. This approach is not without problems because the sequence of NREGA's roll-out was far from random. Khanna and Zimmermann (2017) tackle this using a regression discontinuity design which, while subject to power concerns, finds that NREGA lead to an increase in conflict in the short term.

[3] This contrasts with Blattman et al. (2014), who find that a Ugandan employment program, despite generating large income gains, is not correlated with lower levels of aggression or protests.

They argue that this is driven by increased insurgent repression of civilians, who share more information with security forces in anticipation of the NREGA-led development. This paper steers clear of these concerns by exploiting variation solely within - not across - the NREGA implementation phases, and by focusing on the role of NREGA in providing insurance.

The paper is organized as follows: Section 2 presents the background and data used in this study. Section 3 presents the main empirical approach and the main findings. Sections 4 and 5 present evidence that NREGA provides insurance and that the insurance reduces conflict levels. Section 6 concludes.

2 Background and Data

2.1 Conflict in India

India faces challenges from several insurgency movements. The Maoist insurgency, which is active across many parts of India, is among the most prolific, and it has been a particular focus of the academic literature. The insurgency started as a peasant revolt against landlords and extortive labor practices in the village of Naxalbari in West Bengal in 1967. From there, the "Naxalite" movement (as it is also commonly referred to) spread across India and evolved into an armed insurgency with the stated goals of overthrowing the Indian state and establishing a communist political and social order. In 2004, the two predominant Maoist groups, the Maoist Communist Center (MCC) and the People's War Group (PWG), merged forming the Communist Party of India-Maoist (CPI-M). Estimates of the size of the CPI-M's vary widely, with estimates of anywhere from 5,000 to 40,000 armed cadres, and up to 100,000 village militia members.[4]

The Maoist insurgency is a regional phenomenon, which is clearly visible in the left panel of Figure 1. The map plots districts that form the "red corridor" using a classification from the Indian Ministry of Home Affairs. This coding will be used throughout in the remainder of the paper. Districts in the red corridor are less developed: they have lower rates of urbanization, higher degrees of illiteracy, and limited access to infrastructure (such as paved roads, electricity, primary education and health care facilities, see Table 1). Economic livelihoods in red corridor districts are dominated by subsistence farming, sharecropping, and wage employment in the agricultural sector. As this paper underscores, the income-generating process of the agriculture sector

[4]See e.g. http://goo.gl/xfVewL, accessed 20.01.2013.

strongly depends on rainfall during the monsoon season.[5]

The Maoists are entrenched in rural communities, with a network of village militias through which, for example, Jan-Adalats (people's courts) settle disputes at a local level (Chakrabarty and Kujur, 2009). The Maoists organize riots and protests and, during times of economic duress, launch "famine raids" on grain storage facilities (Dash, 2006). Through the organization of bandhs (strikes), the Maoists push for higher wages in rural areas (Ranjan and Prasad, 2012) and coerce money lenders into relaxing their credit terms during times of distress (Srivastava, 2006). The organization of these types of violent and non-violent activities is often directly related to poor economic conditions, and is used to foster popular support for the armed struggle against the state.

Recruiting of fighters is central to the Maoist efforts to maintain the armed struggle. The Maoists are said to pay their fighters monthly stipends, and to actively recruit teenagers into service. Verma (2011) argues that environments of deprivation following a bad harvest allow the "Maoists to step in, by paying a handsome amount of around 3,000 rupees to the young and promising parents that their kids will have food and money." Accounts suggest that the Maoists pay their fighters monthly stipends of around 1,500 rupees (Ramana, 2007). This figure is significant when compared to average wages for agricultural day laborers, who receive between 50 and 70 rupees per day (in India's poorest districts (see Table 1). This highlights the role that economic shocks to incomes can play both in fostering insurgency activities and recruitment, a crucial link that an effective social insurance can break. The next section discusses how this paper leverages a new conflict event data set built from a large raw text corpus.

2.2 Measurement of Conflict

This paper leverages computational linguistic methods to retrieve conflict event information from a large corpus of 41,347 newspaper clippings gathered by the South Asia Terrorism Portal (SATP) from 2000 to 2014 (see Manning et al., 2008 for an introduction to information retrieval). As the SATP corpus is the most extensive and systematic collection of news coverage of conflict in India, the resulting conflict dataset significantly expands both the temporal and spatial coverage of other currently existing data sets available. Specifically, by covering more than 61 distinct, English-language sources, the corpus vastly expands on the small sets of press agency sources typically

[5]Data from the National Sample Survey 2001 suggest that 64.9% of Indian households rely on agriculture as a main income source. In Chhattisgarh, one of the main Maoist states, up to 90% of the population is employed in agriculture, with generally poor development indicators and higher levels of food insecurity very common in agriculture-intensive areas (Radhakrishna and Ray, 2006).

used in traditional conflict event data bases, such as the Global Terrorism Database (GTD).[6] As a result, the data include nearly 10 times as many conflict events.

To illustrate how conflict event information is extracted from raw texts, consider the example of a sentence below from a typical newspaper clipping. I leverage the SENNA Natural Language Processing package developed by Collobert et al. (2011). Upon parsing, the package provides a part of speech and a parse tree for each sentence, labeling the syntactic role of each word within a sentence. As a result, for every verb, the process identifies its underlying subject, object, and surrounding meta-information (such as time and location). These are indicated by prepositions or their syntactic positions in a sentence. From this starting point, the data undergo further refinement in a sequence of steps. The first step focuses on sentences that include verbs that are

$$\underbrace{\text{Two unidentified terrorists}}_{\text{subject}}\ \overbrace{\text{massacred}}^{\text{verb}}\ \overbrace{\text{six members of a family}}^{\text{object}}$$

$$\text{and}\ \overbrace{\text{left}}^{\text{verb}}\ \overbrace{\text{a seventh}}^{\text{object}}\ \overbrace{\text{injured}}^{\text{verb}}\ \overbrace{\text{at Mangnar Top, Poonch district, on December 31, 2001.}}^{\text{time \& location}}$$

indicative the occurrence of a violent event. The set of verbs considered in this way is broader than would be generated through common manual coding approaches that would only look at subsets of of texts containing certain keywords (such as "to kill").[7] The second refinement standardizes the extracted locational information; this defines a common spatial resolution, which, in the case of the SATP corpus, is the district level.[8] The last step labels actors, and counts conflict casualties. This process involves collecting human judgements from a crowd-sourcing platform, and squaring these human judgements with a trained support-vector machine. The focus is particularly on the object of a verb. In the above example, the object of the verb "massacred" is "six members of a family," which is coded as civilian casualties. The data construction and coding are described in more detail in Appendix B.1, which also benchmarks the dataset against other conflict event data, highlighting that the coverage is much broader.

For the empirical exercises, I focus on two measures of conflict: conflict intensity,

[6] The results are robust to using the Gawande et al. (2017) local language source-based conflict data. Appendix Table B1 lists the most frequently sources, while Appendix Table B2 provides an overview of other papers that have hand-coded smaller subsets of the SATP corpus.

[7] The set of verbs considered is presented in appendix B.1. The approach is similar to what is used in global event data bases such as IECWS (Boschee et al., 2018) or GDELT (Leetaru and Schrodt, 2014).

[8] The populations of all unique locations are matched against a gazetteer of district and location names to achieve a geographic resolution at the 2001 census district level.

captured as the total number of conflict events in a district and year, and a measure of the total number of persons either killed or injured. I focus on the set of 239 districts which have variation over time in both these measures from 2000-2014. Figure 2 plots the total number of conflict events per year captured in my data. Overall, conflict intensity is relatively low, with an average of 1,169 conflict events per year. Conflict has increased over time, most markedly for the poorest districts, which received NREGA early on in Phase 1.

2.3 Productivity shocks and agricultural output measures

Rainfall, especially during the monsoon season, is a central input to the agricultural production process in India. To empirically confirm the relevance of rainfall for incomes, and to study whether NREGA has changed these relationships, I draw on three different data sources. First, I construct a measure of agricultural output using state-level farm harvest prices and annual crop production statistics by district, crop and fiscal year (mapped to the nearest calendar year) from the the Directorate of Economics and Statistics. This yields an unbalanced panel covering the period from 1999 to 2011.[9] Second, I use average annual wages at the district level to construct a measure of labor productivity; I leverage data from the Agricultural Wages in India (AWI) series. This also yields an unbalanced district level panel from 1998 to 2010 [10] Lastly, I draw on data from the Indian Human Development Household Survey (IDHS), for which a first wave was conducted in 2004/2005. A second wave was conducted in 2011/2012, after NREGA had been introduced (Desai and Vanneman, 2018). The survey provides detailed information on the income sources of households over the past year and, for the second wave, asks about NREGA participation.

A central focus of this paper is to study how incomes in rural areas are affected by monsoon season rainfall and explore how these relationships have changed due to NREGA. To measure monsoon season rainfall, I draw in data from the Tropical Rainfall Measuring Mission (TRMM) satellite, which is the highest-quality, remote-sensed rainfall dataset, with global coverage currently available.[11] The monsoon accounts for around 70% of India's annual rainfall and is crucial for agricultural production. Yet, rainfall from the monsoon is not uniform across India. Figure 3 illustrates the typical growing season cycle, the distribution of rainfall and employment under NREGA for

[9] The fiscal year begins in April and ends in March of the subsequent calendar year. The full list of 22 crops and more information about the construction of output value are provided in Appendix B.6.

[10] Berg et al. (2018) is a recent paper using this data. Appendix B.7 presents the data in more detail in discusses some known data issues.

[11] Appendix B.8 provides more details on the data.

the state of Andhra Pradesh across a year. Using state-level crop calendars, I construct a state-specific monsoon period.

This is used to construct district-level monsoon season rainfall by calendar year. I convert the monsoon season rainfall data at the 0.25 x 0.25 grid level to an unweighted district-level average across the grid cells. The resulting measure provides the average amount of rainfall (in millimeters) that a district receives. I also use as a placebo the rainfall outside the monsoon season, which has a much weaker link with incomes. The next section provides background and introduces the data I use to measure NREGA participation.

2.4 NREGA Workfare Program

NREGA, established in 2005, is a legal entitlement to 100 days of (minimum) wage paid public employment per household, per fiscal year, in rural areas. It was rolled out in three phases. A first set of 200 districts received it in 2006; another 130 districts added in 2007; and from 2008 the program was available in all Indian districts (except a few urban centers). The rollout was far from random. Table 1 highlights that districts receiving NREGA early had significantly lower agricultural output per capita, lower wages, and worse access to public goods. They were also more likely to be part of the red corridor and to have experienced higher levels of conflict. This suggests that districts across different NREGA implementation phases are likely poor control groups. As a result, I make sure not to exploit this variation in the empirical design.

NREGA is administered at the village level, and is designed to be demand led. Employment on public works projects is provided once households express their willingness to work, with the village governing body (Gram Panchayat) required to provide the specified number of days of work within a two-week period. The wage rate is fixed at the state-level minimum wage. Villages are empowered through the act, which gives them a say in what local public goods are needed. NREGA public works broadly fall into categories of drought-proofing, micro-irrigation, sanitation, and road construction. While infrastructure developed through NREGA could have an independent effect on the relationship between monsoon rainfall and incomes, no strong evidence of this channel is detectable in the data I study.

In order to study whether NREGA provides insurance, I obtain official participation data for the period between 2006 and 2013. I focus on expenditure per capita, the share of households participating under NREGA in a given fiscal year, and the total number of days worked per capita. I study how these figures respond to variation in local monsoon season rainfall. To be consistent throughout, I match the fiscal year that ranges from April to March to the nearest calendar year.[12]

NREGA participation is most widespread in districts that received NREGA early on (in phases 1 and 2) of the rollout, as highlighted in Panel F of Table 1. The participation rate in these districts is significantly higher compared to the 31% rate of participation in districts that entered the program in the third phase. Similarly, expenditure per capita in these districts is around one-third higher, reflecting a higher baseline demand for NREGA employment. NREGA is available and widely used in areas affected by conflict. The 239 districts that experience some conflict events in my data account for more than 52% of the overall NREGA expenditures across India. As I show, these are also the areas where demand for NREGA is most sensitive to local monsoon variation. Rather than being opposed to the program, the Maoists have in some cases demonstrated outright support for it; for example, they have put up posters urging villagers to claim their rights under the act.[13]

In order to shed light on the extent to which NREGA provides insurance, I study how monsoon season rainfall in the previous growing season increases demand for NREGA employment prior to the next agricultural cycle. Figure 3 illustrates the timing of NREGA employment relative to the monsoon cycle by using the case of the state of Andhra Pradesh in the 2011-2012 period. Take-up is concentrated in the lean season before the start of the new agricultural cycle. The number of households participating peaks in May at around 4.5 million households, and declines to just around 200,000 during the harvest months when agricultural labor demand reaches its peak.

The design of NREGA has several appealing features. NREGA requires that households actually work at minimum wages; thus, it induces self-selection of individuals with poor outside options, and ensures that NREGA assistance is better targeted (see Besley and Coate, 1992; Nichols and Zeckhauser, 1982). Further, infrastructure construction through public employment simplifies monitoring of the program as output is easily verifiable. This can reduce moral hazard problems that could arise due to the decentralized nature of the program. As a result, it has been suggested that NREGA is much better targeted than other welfare programs (see Desai et al., 2015). Yet, NREGA also faces implementation challenges. Early on, leakage of NREGA funds and withholding of wages were documented (Niehaus and Sukhtankar, 2013a,b). Over time, implementation quality has improved, aided by the introduction of a national biomet-

[12]Refer to Appendix B.9 for further discussion of the available NREGA data.

[13]See Hindustan Times, `http://goo.gl/r4g8mm`, accessed 22.04.2013.

ric identification card (Muralidharan et al., 2016), direct transfers of wage payments to recipients' bank accounts, and ICT-fostered improvements in administrative processes (Banerjee et al., 2017). The most-cited barrier that may undermine NREGA's role in providing insurance is the failure of *panchayats* to provide work (Dutta et al., 2014). For the purposes of this paper, the main threat is that rationing itself is more prevalent during periods of economic distress induced by poor monsoon season rainfall. Complementing the analysis of official NREGA data, I draw on self-reported NREGA participation data from the second wave of the IDHS. This allows me to study whether rationing is correlated with monsoon season rainfall, and will reveal that rationing is most prevalent when overall demand for NREGA work is low. This is consistent with the idea that *panchayats* may be reluctant to incur the fixed cost of drawing up public works projects when overall demand for work is low.

The size and, more importantly, the direct delivery of NREGA resources to rural households could have an impact on the relationship between local monsoon shocks and income in the agricultural sector; through this channel the NREGA could have an affect on conflict. I next discuss the empirical strategy and presents the main results.

3 Empirical Strategy and Main Results

3.1 Empirical Strategy

I present results from two sets of main specifications. Specification 1, estimated through OLS, examines how NREGA changes the way that contemporaneous monsoon rainfall in district d and calendar year t, $R_{d,t}$, affects agricultural output or wages, $y_{d,t}$.

$$log(y_{dt}) = \eta \times \log(R_{d,t}) + \theta \times T_{dt} \times \log(R_{d,t}) + a_d + b_{p(d),r(d),t} + \epsilon_{dprt} \quad (1)$$

$$\mathbb{E}(A_{dt}) = a_d \exp\left[\beta \times \log(R_{d,t-1}) + \gamma \times T_{dt} \times \log(R_{d,t-1}) + b_{p(d),r(d),t} + \nu_{dprt}\right] \quad (2)$$

Specification 2 examines how NREGA changes the relationship between lagged monsoon season rainfall $R_{d,t-1}$ and conflict, A_{dt}, measured either as the total number of conflict events or the total number of casualties. Given the count data nature, this specification is estimated using a Poisson model. The indicator T_{dt} is a dummy variable, indicating whether NREGA is available in a district d at a point in time t.

All specifications include district fixed effects, a_d, which absorb any time-invariant characteristics that explain different levels of agricultural output, wages or conflict. Further, the specifications control for non-linear time trends, indicated by $b_{p(d),r(d),t}$. These time effects are specified at the level of the geographic region r by NREGA implementation phase p pair. Hence, the time effects control for shocks that are common to the groups of districts that start to receive NREGA from the same point in time and are located within the same geographic region.[14] As a result, the treatment indicator, T_{dt}, is perfectly collinear with these time effects. This ensures that I do not exploit variation across NREGA implementation phases, which is prudent as the assignment of districts to the three NREGA roll out phases p was not random (see Table 1).

The only difference between specifications 1 and 2 is the timing of the monsoon rainfall measure. The contemporaneous monsoon rainfall measure $\log(R_{d,t})$ affects contemporaneous wages and agricultural output y_{dt}, while it is the lagged monsoon shock $\log(R_{d,t-1})$ that translates into conflict. The choice of empirical design closely follows the income-generating process: the main harvest happens towards the end of the calendar year, which coincides with peak labor demand, see Figure 3. This empirical model also maps well onto the existing literature that documents a lagged effect of (proxies of) income on the intensity of conflict in India (see Gawande et al., 2017; Vanden Eynde, 2016).

The focus of the empirical analysis rests on the interplay between the estimated coefficients η and θ in specification 1, and β and γ in specification 2 respectively. The coefficient η captures the relationship between monsoon rainfall and agricultural output and wages. We would expect these coefficients to be positive, $\eta > 0$, indicating that negative monsoon shocks translate into lower output and wages, respectively. On the other hand, we expect the estimate on the reduced form relationship linking monsoon rainfall and conflict to be negative, $\beta < 0$. The causal chain is simple: less monsoon rainfall translate into lower incomes due to reduced productivity, which increases conflict. The estimates of the coefficients θ and γ capture how these relationships have changed, on average, for the years following the introduction of NREGA. With regard to agricultural wages, we would expect that the estimated coefficient on the NREGA interaction would be negative, $\theta < 0$. NREGA establishes a binding wage floor, which should weaken the pass through of negative shocks, in particular. With regards to agricultural output, the impact of NREGA is ambiguous. The infrastructure constructed through NREGA, such as micro-irrigation, could directly alter the functional relationship between rainfall and output; thus, NREGA could induce farmers to plant crops that are more susceptible to weather shocks, but yield higher expected return (see

[14] Districts d are mapped to three geographic regions $r(d)$. The regions are the North-East, including Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, Sikkim; the states making up the red corridor in the east, (Andhra Pradesh, Bihar, Chhattisgarh, Jharkhand, Karnataka, Maharashtra, Orissa, West Bengal); a third region contains the remaining states in the west. Given that there are three NREGA implementation tranches, there are nine region by implementation phase pairs.

Gehrke, 2017 for a paper presenting evidence in support of this channel).

Lastly, regarding the reduced form specification to study conflict, we expect the estimated coefficient on the interaction to be positive, $\gamma > 0$, relative to the negative coefficient, $\beta < 0$. Of interest are the sign and the joint significance of the estimated coefficients $\beta + \gamma$. If the sum of these two coefficients turns out to be small in magnitude, this suggests that the relationship between monsoon rainfall and conflict has become weaker in the wake of the introduction of NREGA. This is the central hypothesis of this paper.

3.2 Results

I present the main results in Table 2. Column (1) shows how agricultural output per capita is a function of monsoon season rainfall, focusing on the interplay between the coefficient before and after NREGA was introduced. The point estimate suggests that, on average, a 1% reduction in monsoon rainfall decreases output value by around 0.32%. This relationship is broadly unchanged after NREGA has been introduced. This is not surprising: NREGA aims to produce sustainable local infrastructure, which in the longer run could make the agricultural production function less dependent on rain. On the other hand, changes in crop choice could increase the sensitivity as suggested by Gehrke (2017). Thus, the absence of such an effect is unsurprising.

Column (2) presents results for agricultural wages. Before the introduction of NREGA, the association between monsoon rainfall and wages is significant and positive, suggesting that negative rainfall conditions translate into lower wages. Studying the NREGA interaction effect with the monsoon rainfall measure suggests that the relationship between monsoon rainfall and wages has disappeared in the wake of NREGA's introduction. NREGA provides a safe outside option, implying that the program stabilizes incomes for participating households, and has a general equilibrium effect on agricultural labor markets (Muralidharan et al., 2016). This finding extends the existing research on NREGA's impact on agricultural labor markets (Berg et al., 2018; Imbert and Papp, 2015).[15] Turning to conflict outcomes, these results are presented in columns (3) and (4). Before the introduction of NREGA, the elasticity between monsoon rainfall and conflict intensity – as measured by the number of conflict events in column (3) – suggests that a 1% reduction in monsoon rainfall increases

[15] Given that Berg et al. (2018) have pointed out some issues with the AWI data it is reassuring that Appendix Table A1 highlights that similar effects on wages can be found when studying alternative wage data from the village-level component of the two rounds of the IDHS survey. In the table, I show the contrast between the wages of agricultural day laborers and those of skilled workers; the effect on stabilizing wages is driven by agricultural day laborers for whom the NREGA wage floor is likely binding.

conflict intensity by 1.4%. A 25% shock to monsoon rainfall would increase conflict intensity by around 33%. Given that the average district in the estimating sample experiences 4.5 conflict events per year, this would translate into 1.5 additional events per district or about 1006 conflict events. The result for the number of casualties (fatalities and injuries) in column (4) is very similar. This suggests that the relationship between monsoon rainfall and conflict has changed markedly since the introduction of NREGA. The test of the joint significance of the two coefficients fails to reject the null hypothesis of no association between monsoon rainfall and conflict in the wake of NREGA's introduction.

Common Trends A main concern of the analysis is whether the relationship between monsoon rainfall and conflict had already changed prior to the introduction of NREGA (i.e., whether trends in the relationship between monsoon rainfall and conflict are wrongly attributed to NREGA). One such confounder could be India's steady economic growth, which could have lead to a gradual weakening of the relationship between income shocks and conflict. To test for this, I use a more flexible version of specification 2, where I estimate a separate coefficient for how monsoon rainfall is related to conflict for each point in time β_t relative to the NREGA introduction. The estimated specification is:

$$\mathbb{E}(A_{dt}) \quad = \quad a_d \exp\left[\sum_{w=1}^{17} \beta_t \times Time_{p,t} \times log(R_{d,t-1}) + b_{p(d),r(d),t} + \nu_{dprt}\right] \tag{3}$$

The results plotting out the estimated coefficients, $\widehat{\beta_t}$ are presented in Figure 4. The vertical line around zero refers to the period when NREGA was introduced. The solid blue lines indicate the regression coefficients obtained from the baseline pooled specification. The estimated coefficients suggest a consistently negative relationship between monsoon rainfall and conflict before the introduction of NREGA, with the relationship only getting markedly weaker after NREGA was introduced.

Non-parametric analysis In addition to the parametric approach, I also present results from a non-parametric analysis that highlights non-linearities, which may help shed light on the mechanisms that explain the NREGA effect. I follow Hsiang et al. (2013), capturing non-linearities by estimating local linear regressions. The approach incorporates uncertainty in the shape of functional form by estimating these local linear regressions on 1,000 bootstrapped samples. The uncertainty is then visually represented by decreasing the color saturation of the non-parametric regression lines when the variance of estimates increases. I also present results from a more conventional approach to estimate non-linearities using different deciles of the monsoon rainfall variable.

The result from this analysis is presented in Figure 5. The figure plots out bootstrapped lowess estimates of the relationship between monsoon rainfall and the dependent variables for the periods before (Panel A top row) and after NREGA was introduced (Panel B bottom row).[16] The relationship between monsoon rainfall and agricultural output per capita indicates a strong positive relationship: increasing monsoon rainfall increases output, but at a decreasing rate for higher levels of monsoon rainfall. Importantly, however, this relationship looks fairly similar before (top) and after the introduction of NREGA (bottom). Moving to column (2), the relationship between agricultural wages and monsoon rainfall looks quite different in panel A compared to panel B. Before the introduction of NREGA, there is significant pass-through of low monsoon rainfall realizations resulting in lower wages (mapping well onto the findings of Sarsons, 2015; Burgess et al., 2014; Jayachandran, 2006). The non-parametric results suggest that low monsoon rainfall has ceased to depress wages since the introduction of NREGA.[17]

The last column explores the relationship between monsoon rainfall and conflict. Before the introduction of NREGA, there is a weak U-shaped relationship. Places, that experience low monsoon season rainfall see an increase in conflict, while places that see a relatively positive realization above the district-specific mean seeing reductions in conflict, though again, at a decreasing rate as for excessive monsoon rainfall effect, the monsoon rainfall and conflict relationship starts bending back up. This implies that a linear regression, for the period before NREGA, is likely to underestimate the effect of monsoon rainfall on conflict, as the positive tail pushes up the estimate of the slope of the linear regression. Following the introduction of NREGA, the relationship changes fundamentally. Throughout, there is no statistically significant association between monsoon rainfall and conflict. This, together with the linear regression results, suggests that any statistically and economically significant association between monsoon and conflict has disappeared.

[16] Appendix Figure A1 presents very similar results using an alternative approach using rainfall deciles.

[17] In Appendix Figure A2, I show that very similar results can be found using data on unskilled agricultural laborer wages from the the village-level component of the IDHS survey. The results are not inconsistent with Kaur (2018), who finds that real wages do respond to shocks, while nominal wages are downwardly rigid. The region-phase-specific non-linear time trends capture price changes, and hence, the nominal wages are indirectly deflated in my setup.

Robustness checks Appendix Table A2 highlights that the results presented here are not an artefact of the conflict dataset that was leveraged. Using the local-language newspaper based Maoist conflict dataset developed by Gawande et al. (2017) covering the period 2001-2009, across four state and 67 districts as of the 1991 district boundaries, I find very similar results. Appendix Table A3 highlights that results are robust to alternative transformations of the rainfall variable and to using alternative sources of the rainfall data. Appendix Figure A3 shows that results are further robust to dropping each state in turn, alleviating concerns that the results may be driven by any one specific state.[18] I next turn to ruling out alternative explanations.

3.3 Alternative explanations

A main concern is that the observed changes in the relationship between monsoon rainfall and conflict could be driven by other policy changes or, potentially, by changes in police strategy.

Are results due to changes in police strategy? I perform four additional exercises that support the interpretation that the effects documented here are not spuriously driven by changes in police strategy. First, in Table 3, I decompose the conflict-event data across different types of events. Columns (1) and (2) of the table replicate the main results. In columns (3)-(5) I use casualty figures for three different casualty types: civilian, security force and militants. The results suggest that the weakening of the monsoon-conflict relationship is quite uniform, suggesting that the responsiveness of either type of casualties to monsoon season rainfall has not changed in a particularly distinguishable fashion that could indicate a changed police strategy. Column (6) looks at events that report the arrests of militants, while column (7) studies events that do not produce casualties (such as bombings on physical infrastructure). The results are similar to previous results; there is no indication that arrests have become much more prevalent after a bad monsoon shock beyond the wider observation that overall conflict has become less responsive to monsoon variation. This is particularly relevant if there are concerns that the introduction of NREGA may have led to an increased police presence.

Second, in Table 4 I add different sets of additional controls with the baseline results presented in column (1). Indian states are responsible for maintaining law

[18] Results are also robust to alternative functional forms (OLS on the counts, negative binomial or linear probability models). Further results highlight that agricultural output and wages are predominantly driven by monsoon season rainfall (and not by rainfall in the rest of the year or by temperature measures). These are available upon request.

and order; thus, it is reassuring that results are robust to controlling for state by time fixed effects (see column 2). In column (3), I allow for a different set of district fixed effects for the period before and after NREGA was introduced. This absorbs a district-specific NREGA treatment effect on conflict levels. Such an effect on conflict-levels may, for example, account for differential district-level police deployment, which may have a distinct district-specific effect on conflict. Reassuringly, the observed effects on the changing rainfall-conflict elasticity before and after NREGA remain intact. Lastly, column (4) adds a district-specific linear trend, while column (5) uses a less demanding set of time effects. Throughout, the main results remain intact.

Third, I leverage state-level data from the Bureau of Police Research and Development (BPRD). These data, available for the 2004-2014 period, provide information on the size of the state-armed police force, the size of the civil police, and the number of police stations, police outposts (which can be semi-permanent tactical bases), and armed battalions. I add these regressors, varying at the state-by-year level, as controls interacted with the NREGA treatment indicator. This may account for changes in the monsoon rainfall and conflict elasticity, which may be confounded by states expanding their police force or their physical presence across tactical bases in a fashion that may be correlated with the timing of the NREGA rollout. Table 5 highlights that the weakening of the monsoon-rainfall link after NREGA's introduction remains intact even when allowing the time-varying measures of police force strength and the (albeit crude) proxies of their presence to have a heterogenous effect across districts that received NREGA earlier rather than later. Lastly, despite the fact that security and the maintenance of law and order are a state-level responsibility, the central government may assist state security forces through the dispatch of the Central Reserve Police Force (CRPF) for counter-insurgency operations. From early 2010, the CRPF was dispatched due to the launch of the Operation Green Hunt that specifically targeted left-wing extremist-affected districts. The dispatch was coupled with the rollout of the the Integrated Action Plan (IAP), which provided targeted aid spending in areas judged to be affected by the left-wing extremism. Hence, it is not inconceivable that districts receiving IAP funds were also likely to see CRPF deployment. Column (1) of Table 6 shows that allocation of IAP funds was not targeted to areas with a poor monsoon season rainfall. In columns (2) and (5) the sample is restricted to remove the 33 districts that received IAP support. Columns (3) and (6) restrict the analysis to cover the time up to 2010, before much of the IAP activity and Operation Green Hunt started. Lastly, columns (4) and (8) control for IAP spending. Throughout, the observed patterns remain intact, suggesting that the rainfall-conflict elasticity has become markedly weaker since the introduction of NREGA; this allays concerns that results may be confounded by changes in security-force presence or strategy.

Other mechanisms There could yet be a set of alternative plausible mechanisms through which the rainfall and conflict elasticity may have changed since the introduction of NREGA. A first placebo check, which is indicative of the underlying income mechanism, is presented in column (1) of Table 7. Rainfall falling outside the monsoon season (which is a much weaker correlate of agricultural productivity and wages) does not meaningfully correlate with conflict, and NREGA does not change that pattern. This is suggestive evidence that the income-relevance of monsoon season rainfall is driving the association with conflict.

There are several alternative policy shocks that could be plausibly affecting the relationship between income shocks and conflict in a distinct way. In columns (2) and (3) of Table 7, I control for the rollout of the Pradhan Mantri Gram Sadak Yojana Rural Road Construction Scheme (henceforth, PMGSY). The scheme expands rural connectivity and may have a distinct effect on the relationship between income and conflict (see e.g. Burgess and Donaldson, 2010; Fearon and Laitin, 2003). Column (2) indicates that road construction is correlated with a weaker monsoon rainfall-conflict relationship. Once controlling for the NREGA main interaction in column (3) this effect disappears, suggesting that the NREGA interaction better captures the distinct change in the monsoon rainfall and conflict relationship.[19] Another concern is that the timing of the introduction of NREGA coincides with a commodity price boom. Vanden Eynde (2016) finds a weaker effect of monsoon shocks on conflict in places where mining is an important source of income and, thus, household incomes are less affected by weather shocks. In columns (4) and (5), I study the role of the mining sector, as measured by the share of the mining sector in the district domestic product as a cross-sectional characteristic, which is available for 221 out of 239 districts that experience conflict. Out of this sample, 20 report that the mining sector accounts for at least 10% of district domestic product.[20] Column (4) provides evidence that the monsoon rainfall-conflict relationship is weaker in places with a significant mining sector, especially for conflict measured by number of casualties. Again, this effect becomes significantly weaker

[19] These data are not available for all districts that experience conflict. Appendix A.1.1 shows that very similar results are obtained when using different road construction measures.

[20] Appendix A.1.2 provides more detailed discussion of the construction of the sector share variable and of the related results.

once the the NREGA-monsoon rainfall interaction term is included in column (5).

Lastly, columns (6) and (7) also control for the IAP discussed in the previous paragraph. Columns (8) and (9) combine all additional controls. Overall, none of these additional controls that may conceivably account for the change in the rainfall-conflict relationship since NREGA's introduction appear to be absorbing any significant variation. The main coefficients of interest on the monsoon season rainfall and NREGA interaction remain intact.

3.4 Heterogeneity of the NREGA Effect

Given that NREGA was rolled out in three different phases, and that early phase districts were more likely to be part of the Maoist-affected red corridor, I present two heterogeneity exercises along these two dimensions. The results are presented in Table 8. Panel A presents the baseline results for reference. Panel B explores the heterogeneity by NREGA implementation phase. Column (1) in panel B suggests that the poorest phase 1 and 2 districts drive the moderation in the monsoon rainfall-conflict relationship. This is not surprising; these are the districts in which the relationship was initially most pronounced.[21] This maps well onto the findings of the existing literature, which suggests that the economic effects of NREGA are concentrated in districts that received NREGA early (Imbert and Papp, 2015).

Panel C explores the effect inside and outside the red corridor (which is plotted on the left panel of Figure 1). The results suggest, that the moderation in the monsoon rainfall and conflict relationship is coming from these districts. Around 54.7% of the conflict events take place in these districts; this maps well onto the findings of Vanden Eynde (2016) and Gawande et al. (2017).

4 Does NREGA Provide Insurance?

I next study whether and to what extent NREGA participation responds to adverse monsoon seasons. To do this, I leverage both administrative and survey-based data.

Monsoon-induced NREGA take-up The empirical specification used to study NREGA participation follows closely the approach outlined in the previous sections:

$$P_{dt} = \delta_{d_k} + b_{p(d),r(d),t} + \phi \times log(R_{d,t-1}) + \epsilon_{dt} \tag{4}$$

[21] Appendix Figures A4 and A5 present the results graphically, highlighting that common trends also hold when considering NREGA implementation phase and when contrasting districts in the red corridor.

where P_{dt} measures participation levels in district d in year t. I focus on three measures of participation: the total expenditure per capita, the number of days worked per person, and the share of households that demand employment in a district.[22] As before, I control for region by NREGA phase-specific non-linear time effects and district fixed effects.[23] The timing of the monsoon measure is lagged and thus follows the approach in the previous sections. This is the natural approach given that the bulk of NREGA employment falls into the lean season (see Figure 3). The coefficient of interest is ϕ. We would expect a negative coefficient, $\phi < 0$, indicating that a good monsoon in the last agricultural cycle (i.e. high income), is correlated with low demand for NREGA. A negative coefficient is thus suggestive evidence that households rely on NREGA employment as a form of insurance.

Results from this analysis are presented in Table 9. Panel A presents the baseline results, exploring three margins of take-up: expenditure per capita, number of days worked per capita, and the share of households participating. Throughout, the indicators suggest that measures of NREGA take-up strongly counteract monsoon shocks. There is some indication, albeit not across all estimates and outcomes, that the relationship is more precisely estimated when studying districts that received NREGA funds in phase 1 (see Panel B), and/or districts that are part of the red corridor (see Panel C). This is not inconsistent with the previous results that suggested that the moderation in the monsoon rainfall-and-conflict relationship is driven by these districts.

Appendix Figure A6 studies non-linearities: demand for NREGA work is increasing following a low monsoon rainfall, but also following excessive monsoon rainfall, further highlighting the role that NREGA may play in providing insurance. Appendix Table A7 studies survey-based micro data from the second wave of the IDHS household panel. It shows a similar relationship, suggesting increased participation following a bad monsoon.[24]

[22] The dependent variable here is not transformed in logs as in some district-years there is zero demand. The estimated coefficients can be interpreted as semi-elasticities.

[23] The only difference is that I allow the district fixed effects d_k to change discretely once from 2010 onward. This is necessary because NREGA data recording changes discretely, which generates district-specific jumps. This includes variable redefinitions and other public programs that are subsumed under NREGA and are specific to districts. Appendix B.9 provides further detail.

[24] Further, Appendix Table A7 suggests that, while rationing of NREGA is prevalent, the extent of rationing is not correlated with monsoon rainfall. Rather, rationing is most pronounced during periods when overall demand for NREGA work is low (i.e., following a good monsoon). This is could be driven by the fact that drawing up NREGA public works projects involves fixed costs that panchayats may not be willing to incur if overall demand is low.

How much insurance does NREGA provide? To study this question, I explore how total NREGA expenditure per capita P_{dt} moves with variation in lagged agricultural output value per capita, $y_{d,t-1}$, with the latter instrumented by the lagged monsoon shock $log(R_{d,t-1})$.

$$P_{dt} = \delta_{d_k} + b_{p(d),r(d),t} + \xi \times \widehat{y_{d,t-1}} + \nu_{dt} \tag{5}$$

I expect that the estimated coefficient ξ in specification 5 is negative, $\xi < 0$, indicating that high levels of agricultural output are associated with low NREGA expenditures. Since the scales are identical, there is a natural interpretation for the estimated coefficient ξ: it captures (in rupees) the increase in NREGA expenditures in a given district for every one-rupee loss in agricultural output due to a bad monsoon. The results from this simple quantification exercise are presented in Table 10. Column (1) shows how monsoon rainfall is associated with agricultural output per capita. The coefficient on monsoon rainfall shows a semi-elasticity, indicating that a 10% increase in the monsoon rainfall increases the nominal agricultural-output value by 89.7 rupees. This can be interpreted as the first stage of the analysis. The coefficient in column (2) is as in the previous table. In column (3), I present the results from an instrumental variables exercise. Here, lagged agricultural output is instrumented with lagged monsoon rainfall. The coefficient suggests that a 100-rupee loss in agricultural output due to monsoon variation translates into an increased NREGA expenditure of 19.6 rupees. Though this is not full insurance, it suggests that a significant share of agricultural output losses - around 20% - is offset by NREGA resources flowing into a district.

This estimate is likely a lower bound for several reasons. First, the estimate captures the direct financial offset in the rainfall-induced reduction in *agricultural output value per capita*, which is compensated by NREGA transfers flowing into a district. Yet, individuals benefiting from NREGA most likely are neither full residual claimants of the output nor full residual claimants of output losses (see Banerjee et al., 2002). Further, the output value should capture multiple factor payments, such as rental payment for land, payments for agricultural labor used in the production of the output, and other input costs, such as fertilizers and seeds. This means that expressing the insurance value as output losses compensated by inflows of NREGA resources does not say a lot about the actual insurance value to individual households. Third, as I have previously shown, the outside option offered by NREGA can provide insurance and can make household incomes less rainfall elastic even without direct program participation as NREGA establishes a rigid minimum-wage floor. I next exploit a two-period household panel that provides further evidence that household incomes of agricultural wage workers in rural areas have become much less dependent on rainfall since the introduction of NREGA.

NREGA's impact at the household level Using data from both rounds of the Indian Human Development Survey (IDHS), I study to what extent household incomes in rural areas are rainfall elastic and to what extent this elasticity has changed between the first and second survey waves (in between which NREGA was introduced). I estimate very similar specifications to those presented in the main results on agricultural output and wages, except that the observation is now an individual household i.

The results are presented in Table 11. The first three columns present results studying the whole population of survey participants in rural areas in both IDHS waves. For the population at large, household incomes are rainfall elastic and continue to be rainfall elastic between the two survey waves; a 1% decrease in monsoon rainfall being associated with 0.16% lower household incomes. Columns (3)-(6) focus on the subset of households whose main source of income is agricultural labor. For this subpopulation, a 1% reduction in monsoon rainfall is associated with a 0.26% drop in household incomes before NREGA was introduced; by contrast, after the introduction of NREGA the pass through of monsoon rainfall to wages becomes significantly weaker. This highlights that, while the income-generating process for the public at large remains a function of monsoon rainfall, the monsoon-rainfall induced income risk for agricultural workers has been significantly reduced. This is particularly relevant because agricultural workers are among the households with poorest outside options, earning an average 40,994 rupees per year (around USD 560) in the IDHS sample. In comparison, the average household income for the rest of the sample is 92,191 rupees (around USD 1,261), more than twice as high. As NREGA de facto establishes a rigid lower wage floor, the finding that the poorest households with the worst outside options are most likely to benefit from this wage floor is unsurprising.

The analysis suggests that incomes of agricultural day laborers has become much less rainfall elastic. The NREGA benefit of income stabilization operates both directly - through NREGA participation - and indirectly - through the program's general equilibrium effects on rural labor markets (as previously documented). The next section presents results suggesting that NREGA has caused a reduction in conflict levels in districts in which NREGA provides insurance against monsoon shocks.

5 Does NREGA's insurance reduce conflict levels?

5.1 Empirical Strategy

I next provide more direct evidence to shed light on the question of whether NREGA's provision of insurance against monsoon rainfall-induced income variation contributes to lower levels of conflict. To do so, I construct a cross-sectional measure capturing whether demand for NREGA work is decreasing in the extent of monsoon rainfall. I then examine whether this measure is associated with lower levels of conflict following the introduction of NREGA. To arrive at such a district-specific measure, I estimate a heterogenous effects version of the NREGA participation specification equation 6, allowing for a separate coefficient ϕ_d for every district:

$$P_{dt} = \delta_{d_k} + b_{p(d),r(d),t} + \sum_{d} \phi_d \times log(R_{d,t-1}) + \epsilon_{dt} \tag{6}$$

The coefficients ϕ_d capture the responsiveness of demand for NREGA employment with respect to monsoon rainfall in each district d. I estimate the above specification and retain both the point estimate $\widehat{\phi_d}$ as well as the estimated standard error $\widehat{s.e.(\widehat{\phi_d})}$, computing the ratio $\widehat{T_d} = \frac{\widehat{\phi_d}}{\widehat{s.e.(\widehat{\phi_d})}}$. I then convert the variable into a dummy variable $\mathbb{1}(\widehat{\phi_d})$ taking the value equal to one, if the the $\widehat{T_d}$ is in the lower 20th percentile of the *empirical distribution* among all the estimates $\widehat{T_d}$. This incorporates the degree of uncertainty or precision of different $\widehat{\phi_d}$ estimates, but does not involve a comparison with some hypothetical statistical distribution.[25] As before, I explore three measures of NREGA participation: expenditure per capita, days worked per capita, and the share of households demanding employment to construct the $\mathbb{1}(\widehat{\phi_d})$. I use this cross-sectional measure for a difference-in-differences estimation, comparing conflict levels in districts in which NREGA counteracts monsoon rainfall-induced income variation to those in which this insurance effect was not statistically discernible. I estimate:

$$\mathbb{E}(A_{dt}) = a_d \exp\left[\psi \times T_{dt} \times \mathbb{1}(\widehat{\phi_d}) + b_{p(d),r(d),t} + \omega_{dt}\right] \tag{7}$$

As before, A_{dt} indicates the number of conflict events or casualties in a district and calendar year. The coefficient of interest is ψ; a negative coefficient, $\psi < 0$, indicates that districts in which NREGA cushions monsoon shocks experience a relative drop in conflict levels compared to places in which NREGA does not have this insurance function. The specification includes time effects specific to the NREGA implementation

[25] To be precise, since the estimated values $\hat{\phi}_d$ are, on average, negative indicating that demand for NREGA is deceasing in rainfall, I flip the sign for ease of interpretation.

phase, which, as before, ensures that I do not exploit any variation across NREGA implementation phases to estimate ψ. The main concern for this analysis is whether common trends hold. I show that places in which demand for NREGA is responsive to local monsoon rainfall only start experiencing lower levels of conflict reliably and in a timely fashion around the introduction of NREGA. This provides compelling evidence in support of the common trends assumption.

5.2 Results

In Table 12, I present results from the difference-in-differences analysis, which exploits heterogeneity across districts in the extent to which households' demand for NREGA is a decreasing function of monsoon rainfall. Panel A presents the overall effect across all districts. Throughout, the coefficient is negative and, in most specifications, significantly different from zero. This suggests that conflict levels go down in places in which demand for NREGA increases following a low monsoon rainfall. The size of the coefficient is economically meaningful: across the specifications in Panel A, the coefficient hovers between -0.6 to -0.8, indicating that the difference in the logs of expected counts is expected to be 0.6 to 0.8 log points lower in districts in which NREGA participation offsets local monsoon shocks. This translates to a significant, 40%-50% drop in conflict levels. Panel B explores this effect by NREGA-implementation phase. Again, though estimated with less precision, the results appear to be more concentrated in districts that received NREGA in Phase 1. Lastly, Panel C explores the heterogeneity of the effect, comparing districts forming the red corridor, with districts outside the red corridor. Again, the results suggest that those districts experience drops in conflict levels. The estimated coefficient suggests that the difference in the logs of expected counts of conflict is expected to be one log point lower for districts inside the red corridor, where demand for NREGA is responsive to local monsoon variation. This translates into a relative drop in conflict levels of around 60%.

Robustness Appendix Table A9 presents the results obtained when coding the dummy to take the value of one, in case the estimated coefficient ϕ_d is in the lower 10th or 30th percentile distribution of the estimated $\widehat{T_d}$. The results are very similar to the main results. In Appendix Table A10, I use the level $\widehat{\phi_d}$ and weigh observations by the inverse of the estimated standard error $\widehat{s.e.(\widehat{\phi_d})}$. The results are similar to those obtained with the dummified measure.

Common trends The natural concern with this analysis is that districts in which demand for NREGA responds to monsoon rainfall may have been on differential trends before NREGA was introduced. Figure 6 suggests that this is not likely to be the case. Conflict falls in districts where NREGA demand is responding to monsoon rainfall, but only after NREGA was introduced, suggesting that these places have not been on differential trends in conflict prior to the introduction of NREGA.

Where does NREGA provide insurance? The degree to which ϕ_d is negative may depend on many factors. In order to shed some light onto what drives variation in the ϕ_d, I use a a hands-off machine-learning approach called best subset selection to help identify the vector variables that do a good job in characterizing what drives the variation in the measure $\mathbb{1}(\widehat{\phi}_d)$. The best subset selection (BSS) method I employ finds the solution to the following non-convex combinatorial optimization problem:

$$min_\beta \sum_d (\mathbb{1}(\widehat{\phi}_d) - \beta_0 - \sum_{j=1}^{p} x_{dj}\beta_j)^2 \text{ subject to } \sum_{j=1}^{p} \mathbf{I}(\beta_j \neq 0) \leq s, \tag{8}$$

where p is the set of regressors of which a subset s is chosen to maximize overall model fit. The idea is to find the best model among all potential models to identify the vector of covariates that does the best job in capturing variation in the $\mathbb{1}(\widehat{\phi}_d)$. The result is a sequence of models $\mathcal{M}_1, ..., \mathcal{M}_s, .., \mathcal{M}_p$, where $\mathcal{M}_s$ indicates the best model among the class of models including exactly s covariates. The overall optimal model among the sequence of best models, $\mathcal{M}_{s*}$, is chosen by using either cross validation or some degree-of-freedom-adjusted measure of goodness of fit such as the Akaike information criterion (AIC).[26] I consider a rich set of 31 cross-sectional district characteristics for BSS to chose from. This includes the full set of baseline characteristics from the 2001 Census explored in the summary statistics Table 1; further, I include the NREGA treatment tranche indicator, whether a district is part of the red corridor, and district-specific measures that capture the degree to which agricultural output or agricultural wages are a function of rainfall *prior to the introduction of NREGA*. These are constructed in a similar fashion as the ϕ_d's. Lastly, I also include measures of the share of a district area that has lit pixels using night lights data and whether a district would be selected to be receiving special IAP assistance funding.

The decomposition of what drives the variation across the $\mathbb{1}(\widehat{\phi}_d)$ is presented in

[26] An alternative method would be to use Lasso, which is more widely known and can handle problems when the set of covariates to consider is large; the main drawback is that BSS becomes computationally infeasible very fast (with more than 40 covariates).

Appendix Table A8. I present the first twelve "best" models $\mathcal{M}_1, ..., \mathcal{M}_{12}$. The twelfth model is the one that, using the AIC criterion, is judged to perform best among all feasible models in navigating the bias-variance trade-off. Some of the district characteristics that seem to explain the variation in the $\mathbb{1}(\widehat{\phi}_d)$ are quite intuitive. In particular, both measures capturing the extent to which agricultural wages or agricultural output per capita are a function of monsoon rainfall prior to NREGA's introduction are increasing the likelihood that a district's NREGA demand appears responsive to monsoon rainfall. This is very intuitive, as one would expect that the insurance value of NREGA is largest in places where the income-generating process is most affected by monsoon rainfall. Further, the results suggest that the dummy variable indicating whether a district is inside the red corridor is, not surprisingly, also a characteristic that seems to matter. Some other census characteristics are more difficult to interpret. The tribal population share, along with other measures that may be considered to capture low baseline level of development (such as low prevalence of primary school facilities, a high gender gap or high illiteracy rates), are correlated with the $\mathbb{1}(\widehat{\phi}_d)$ indicator capturing whether NREGA participation appears responsive to monsoon variation.

The results thus indicate that NREGA does provide insurance in the places where demand for that insurance is likely highest. Further, a measure that reasonably captures whether NREGA demand is responding to monsoon rainfall is significantly associated with lower levels of conflict, after the introduction of NREGA. This is further evidence that the observed reduced-form weakening of the monsoon season rainfall-conflict relationship and ensuing lower levels of conflict are due to NREGA's provision of insurance against adverse income shocks.

6 Conclusion

This paper has studied the impact of a social insurance program on conflict in India. The existing conflict literature has devised various identification strategies to exploit exogenous variations in incomes to study the relationship between income and conflict. The findings of this literature have a direct policy implication: any measure that helps insulate household incomes from adverse shocks should moderate the relationship between these exogenous productivity shocks and conflict.

The key findings of this paper suggest that a large scale public employment program in India has eliminated the previously existing relationship between monsoon rainfall and conflict. While weather shocks continue to affect rural areas, these shocks have ceased to translate into conflict as a result of having the public employment program in place. The insurance value delivered by the public employment scheme is significant. A simple quantification exercise suggests that at least 20% of district-level income losses due to adverse monsoon conditions are directly offset by the increased expenditures the program provides. I show that districts where NREGA offsets monsoon-induced income losses experience significant drops in conflict levels.

The paper has important implications for policy makers. Incomes in developing countries are much more volatile, leaving households exposed to significantly more risk than those in developed countries. Climate models suggest that erratic weather events could become even more severe and frequent. This has led to concerns about conflicts in the future (see Hsiang et al., 2013). Many developing countries have not been able to devise policies to provide adequate protection. This paper highlights that social insurance in form of a public employment program reduces conflict, thus delivering social benefits that go beyond mere insurance value.

References

Aggarwal, S. (2014). Do Rural Roads Create Pathways Out of Poverty? Evidence from India. *mimeo*.

Annan, J. and C. Blattman (2016). Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State. *American Political Science Review 110*(01), 1–17.

Banerjee, A., E. Duflo, C. Imbert, S. Mathew, and R. Pande (2017). E-governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India . *NBER Working Paper*.

Banerjee, A. V., P. J. Gertler, M. Ghatak, S. Journal, and N. April (2002). Empowerment and Efficiency : Tenancy Reform in West Bengal. *Journal of Political Economy 110*(2), 239–280.

Bazzi, S. and C. Blattman (2014). Economic Shocks and Conflict: Evidence from Commodity Prices. *American Economic Journal: Macroeconomics 6*(4), 1–38.

Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Berg, E., S. Bhattacharyya, D. Rajasekhar, and R. Manjula (2018). Can public works increase equilibrium wages? Evidence from India's National Rural Employment Guarantee. *World Development 103*, 239–254.

Berman, E., J. H. Felter, J. N. Shapiro, and E. Troland (2013). Modest, secure, and informed: Successful development in conflict zones. *American Economic Review 103*(3), 512–517.

Berman, E., J. Shapiro, and J. Felter (2011). Can hearts and minds be bought? The economics of counterinsurgency in Iraq. *Journal of Political Economy 119*(4), 766 – 819.

Besley, T. and S. Coate (1992). Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs. *American Economic Review 82*(1), 249–261.

Besley, T. and T. Persson (2008). The Incidence of Civil War: Theory and Evidence. *NBER Working Paper*.

Blattman, C., N. Fiala, and S. Martinez (2014). Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda. *The Quarterly Journal of Economics 129*(2), 697–752.

Blattman, C. and E. Miguel (2010). Civil War. *Journal of Economic Literature 48*(1), 3–57.

Boschee, E., J. Lautenschlager, S. O'Brien, S. Shellman, and J. Starz (2018). ICEWS Automated Daily Event Data.

Brückner, M. and A. Ciccone (2011). Rain and the Democratic Window of Opportunity. *Econometrica 79*(3), 923–947.

Buhaug, H. and G. Wischnath (2014). Rice or riots: On food production and conflict severity across India. *in Press, Political Geography*.

Burgess, R., O. Deschenes, D. Donaldson, and M. Greenstone (2014). The Unequal Effects of Weather and Climate Change: Evidence from Mortality in India. *mimeo*.

Burgess, R. and D. Donaldson (2010). Can Openness Mitigate the Effects of Weather Shocks? Evidence from India's Famine Era. *American Economic Review: Papers & Proceedings 100*(2), 449–453.

Burnicki, A. C., D. G. Brown, and P. Goovaerts (2007). Simulating error propagation in land-cover change analysis: The implications of temporal dependence. *Computers, Environment and Urban Systems 31*(3), 282–302.

Chakrabarty, B. and R. K. Kujur (2009). *Maoism in India: Reincarnation of Ultra-Left Wing Extremism in the Twenty-First Century*. Delhi: Routledge.

Chassang, S. (2009). Economic Shocks and Civil War. *Quarterly Journal of Political Science 4*(3), 211–228.

Ciccone, A. (2011). Economic Shocks and Civil Conflict: A Comment. *American Economic Journal: Applied Economics 3*(4), 215–227.

Collier, P. and A. Hoeffler (1998). On economic causes of civil war. *Oxford Economic Papers 50*(4), 563–573.

Collier, P. and A. Hoeffler (2004). Greed and grievance in civil war. *Oxford Economic Papers 56*(4), 563–595.

Collobert, R., J. Weston, and L. Bottou (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research 12*, 2461–2505.

Commission on Fragility and Development (2018). Escaping the fragility trap. Technical Report April.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Crost, B. and J. Felter (2019). Export Crops and Civil Conflict. *forthcoming, Journal of the European Economics Association*.

Crost, B. and P. Johnston (2014). Aid under fire: Development projects and civil conflict. *American Economic Review 104*(6), 1833–1856.

Dal Bó, E. and P. Dal Bó (2011). Workers, Warriors, and Criminals: Social Conflict in General Equilibrium. *Journal of the European Economic Association 9*(4), 646–677.

Dasgupta, A., K. Gawande, and D. Kapur (2017). (When) Do Antipoverty Programs Reduce Violence? India's Rural Employment Guarantee and Maoist Conflict. *International Organization 71*(03), 605–632.

Dash, S. P. (2006). *Naxal Movement and State Power: With Special Reference of Orissa* (1st editio ed.). New Delhi: Sarup & Sons.

Desai, S. and R. Vanneman (2018). India Human Development Survey-II (IHDS-II), 2011-12.

Desai, S., P. Vashishtha, and O. Joshi (2015). Mahatma Gandhi National Rural Employment Guarantee Act: A Catalyst for Rural Transformation. *IHDS Report*.

Dube, O. and J. F. Vargas (2013). Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *The Review of Economic Studies 80*(4), 1384–1421.

Dutta, P., R. Murgai, M. Ravallion, and D. V. D. Walle (2014). *Right to Work?: Assessing India's Employment Guarantee Scheme in Bihar*.

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review 97*(01), 75.

Ferrara, E. L. and M. Harari (2018). Conflict, Climate, and Cells: A Disaggregated Analysis. *The Review of Economics and Statistics 100*(4), 594–608.

Fetzer, T. and S. Kyburz (2018). Cohesive Institutions and Political Violence. *CAGE Working Paper*.

Gawande, K., D. Kapur, and S. Satyanath (2017). Renewable Natural Resource Shocks and Conflict Intensity. *Journal of Conflict Resolution 61*(1), 140–172.

Gehrke, E. (2017). An Employment Guarantee as Risk Insurance? Assessing the Effects of the NREGS on Agricultural Production Decisions. *The World Bank Economic Review* (March).

Gomes, J. (2012). The Political Economy of the Maoist Conflict in India: An Empirical Analysis. *mimeo*.

Guardado, J. and S. Pennings (2017). The Seasonality of Conflict. *mimeo*, 1–38.

Hegre, H. and N. Sambanis (2006). Sensitivity Analysis of Empirical Results on Civil War Onset. *Journal of Conflict Resolution 50*(4), 508–535.

Hoelscher, K., J. Miklian, and K. C. Vadlamannati (2012). Hearts and mines: A district-level analysis of the Maoist conflict in India. *International Area Studies Review 15*(2), 141–160.

Hsiang, S. M., M. Burke, and E. Miguel (2013). Quantifying the influence of climate on human conflict. *Science 341*(6151), 1–7.

Imbert, C. and J. Papp (2015). Labor Market Effects of Social Programs: Evidence from India's Employment Guarantee. *American Economic Journal: Applied Economics 7*(2), 233–263.

Iyengar, R., J. Monten, and M. Hanson (2011). Building Peace: The Impact of Aid on the Labor Market for Insurgents. *NBER Working Paper 17297*.

Iyer, L. and P. Topalova (2014). Poverty and Crime : Evidence from Rainfall and Trade Shocks in India. *Harvard Business School Working Paper 14067*.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy 114*(3), 538–575.

Katyal, R., M. Sardana, and D. Satyanarayana (2001). *Estimates of District Domestic Product*. New Delhi: Socio-Economic Research Centre.

Kaur, S. (2018). Nominal wage rigidity in village labor markets. *forthcoming, American Economic Review*.

Khanna, G. and L. Zimmermann (2017). Guns and butter? Fighting violence with the promise of development. *Journal of Development Economics 124*, 120–141.

Kudamatsu, M., T. Persson, and D. Strömberg (2014). Weather and Infant Mortality in Africa. *mimeo*.

Leetaru, K. and P. A. Schrodt (2014). GDELT: Global data on events, location, and tone. *International Studies Association meetings*.

Leung, Y., Y. Ge, J. Ma, and J. Wang (2005). Measurement Errors and their Propagation in the Registration of Remote Sensing Images. In *Developments in Spatial Data Handling*, pp. 285–297. Springer Berlin Heidelberg.

Li, X.-H., Q. Zhang, and C.-Y. Xu (2012). Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. *Journal of Hydrology 426-427*, 28–38.

Manning, C. D., P. Raghavan, and H. Schütze (2008). *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press.

Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy 112*(4), 725–753.

Ministry of Rural Development (2009). Joint Convergence Guideline. Technical report, Ministry of Rural Development.

Morgan, R. and D. Reiter (2013). How War Makes the State: Insurgency, External Threat, and Road Construction in India. *mimeo*.

Muralidharan, K., P. Niehaus, and S. Sukhtankar (2016). Building state capacity: Evidence from biometric smartcards in Indi. *American Economic Review 106*(10), 2895–2929.

Muralidharan, K., P. Niehaus, and S. Sukhtankar (2017). General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence from India. *NBER Working Paper*, 1–76.

Nichols, A. and R. Zeckhauser (1982). Targeting Transfers through Restrictions on Recipients. *American Economic Review: Papers & Proceedings 72*(2), 372–377.

Niehaus, P. and S. Sukhtankar (2013a). Corruption Dynamics: The Golden Goose Effect. *American Economic Journal: Economic Policy 5*(4), 230–269.

Niehaus, P. and S. Sukhtankar (2013b). The marginal rate of corruption in public programs: Evidence from India. *Journal of Public Economics 104*, 52–64.

Nunn, N. and N. Qian (2014). US food aid and civil conflict. *American Economic Review 104*(6), 1630–1666.

Radhakrishna, R. and S. Ray (Eds.) (2006). *Handbook of Poverty in India: Perspectives, Policies, and Programmes*. Oxford: Oxford University Press.

Ramana, P. V. (2007). *The Naxal Challenge: Causes, Linkages, and Policy Options* (1st editio ed.). Chennai: Pearson.

Ranjan, R. and V. Prasad (2012). Exploring Social Resilience in State Fragility: A Climate Change Perspective. In M. Hamza and C. Corendea (Eds.), *Climate Change and Fragile States: Rethinking Adaptation*, Number 16, Chapter 2. Bonn: United Nations University.

Sarsons, H. (2015). Rainfall and conflict: A cautionary tale. *Journal of Development Economics 115*, 62–72.

Schneider, U., A. Becker, P. Finger, B. Meyer-Christoffer, Anja; Rudolf, and M. Ziese (2011). GPCC Full Data Reanalysis Version 6.0 at 0.5 Degree: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data.

Sekhri, S. and A. Storeygard (2014). Dowry deaths: Response to weather variability in India. *Journal of Development Economics 111*, 212–223.

Shapiro, J. N., M. Callen, E. Berman, and J. Felter (2011). Do Working Men Rebel? Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines. *Journal of Conflict Resolution 55*(4), 496–528.

Shrivastava, A. (2014). Civil conflict with rising wages and increasing state capacity: The Maoist insurgency in India. *mimeo*.

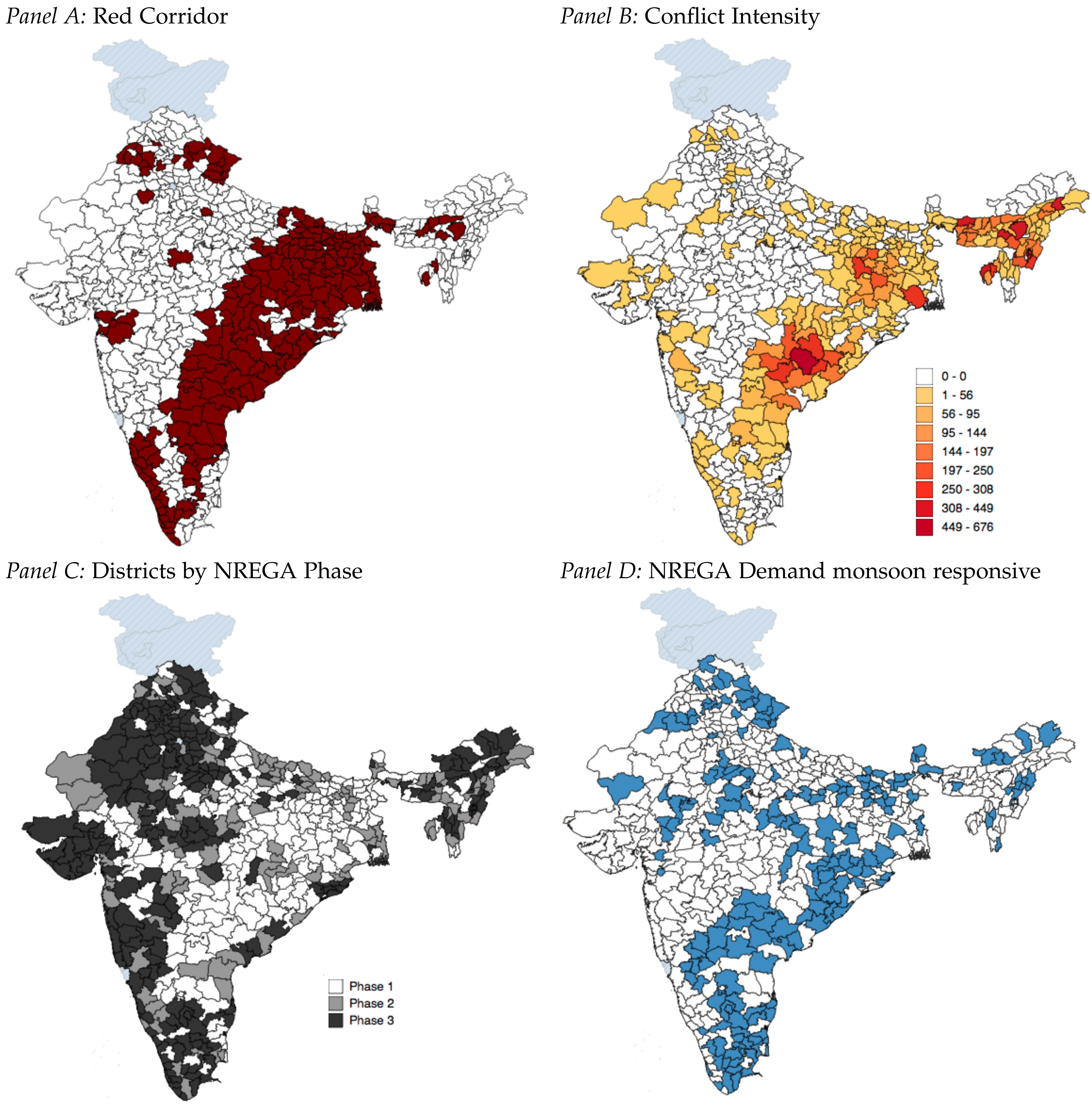
Smith, S. R., M. A. Bourassa, and M. Long (2011). Pirate attacks affect Indian Ocean climate research. *Eos, Transactions American Geophysical Union 92*(27), 225.

Srivastava, A. (2006). *Maoism in India*. Delhi: Prabhat Prakashan.

Vanden Eynde, O. (2016). Targets of Violence: Evidence from India's Naxalite Conflict. *The Economic Journal* (November), 1–50.

Verma, S. (2011). Far Reaching Consequences of the Naxalite Problem in India. Technical Report July, Rakshak Foundation.

Figure 1: Maps of districts across India



Notes: Panel A plots out districts in the "red corridor" according to Government of India, Panel B presents conflict intensity measured by total number of events in the SATP corpus, Panel C highlights the different phases of NREGA roll out, while Panel D plots out districts where demand for NREGA is estimated to be responsive to local monsoon variation.

Figure 2: Total Number of Conflict events by NREGA implementation phase.

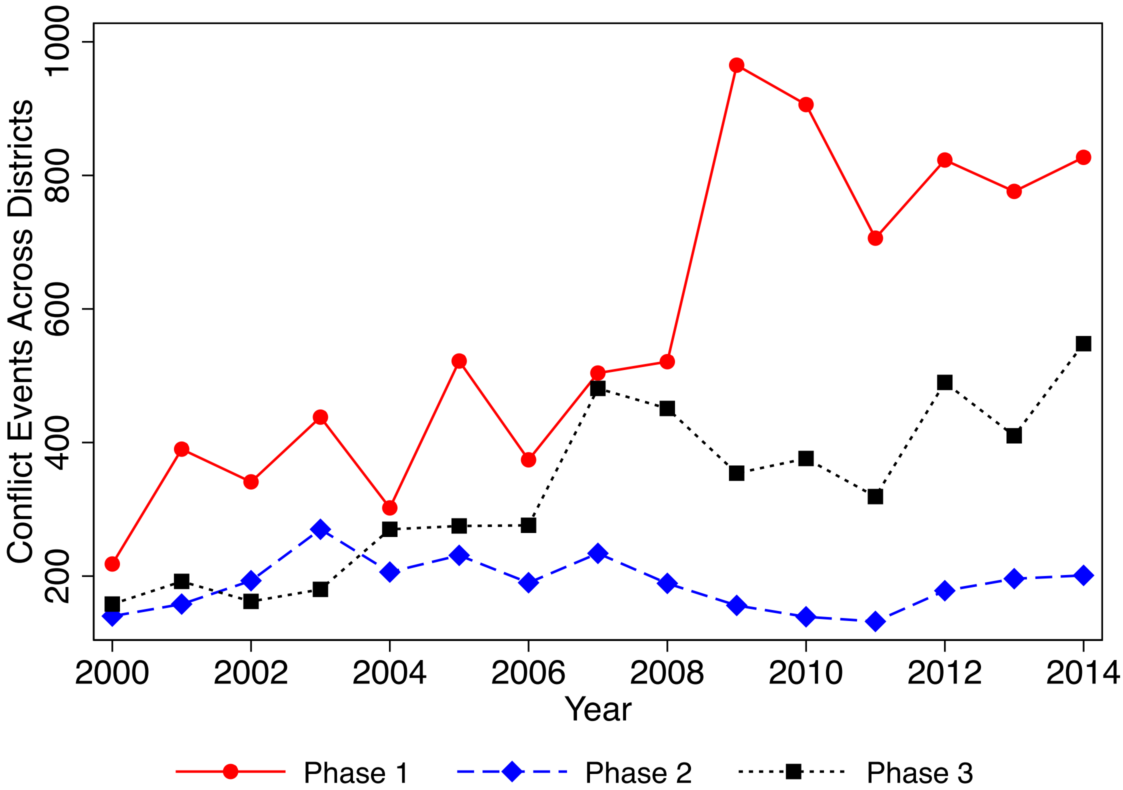


Figure 3: Rainfall (dashed), growing season and NREGA employment (solid line) in 2011-2012 by month for Andhra Pradesh

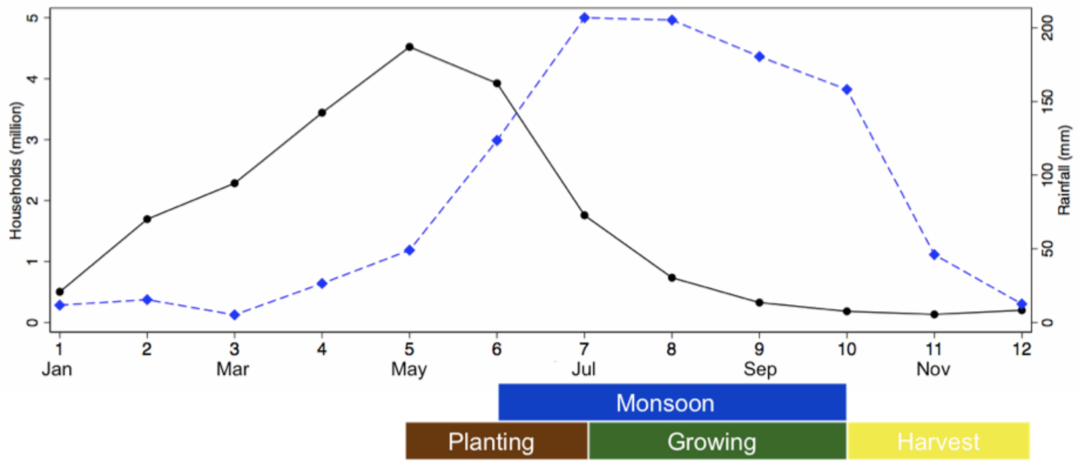
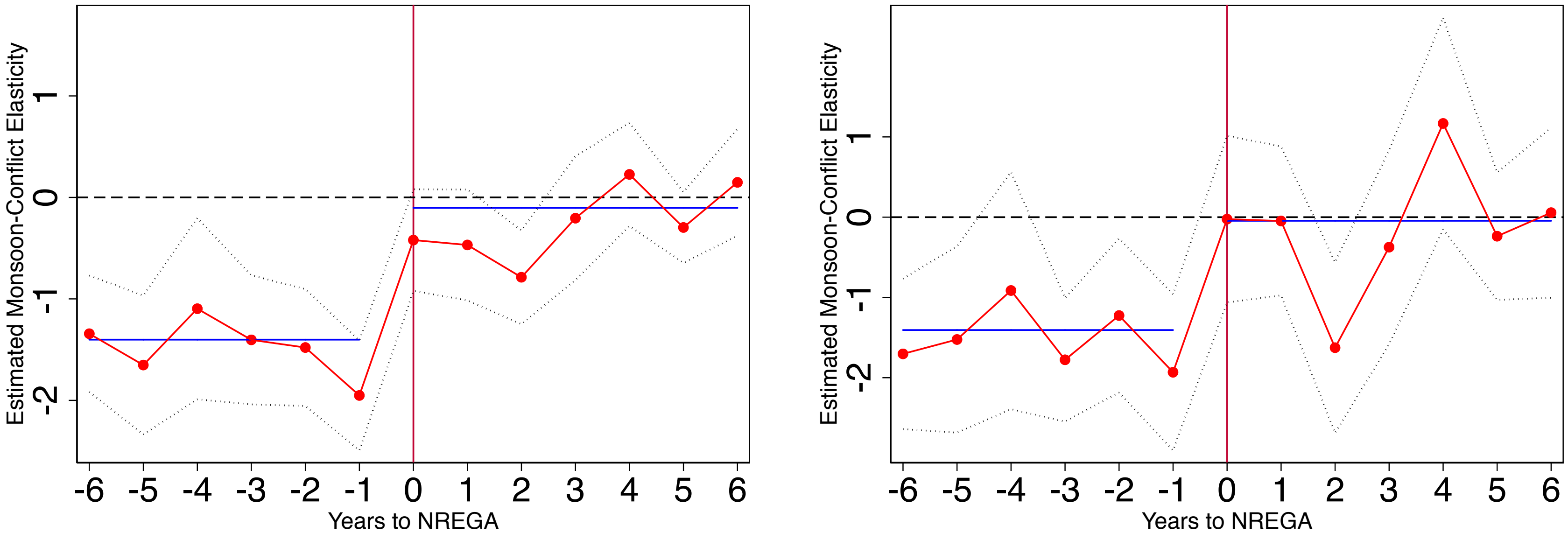


Figure 4: Effect of monsoon rainfall on conflict over time relative to the NREGA introduction

Panel A: Total Number of Events

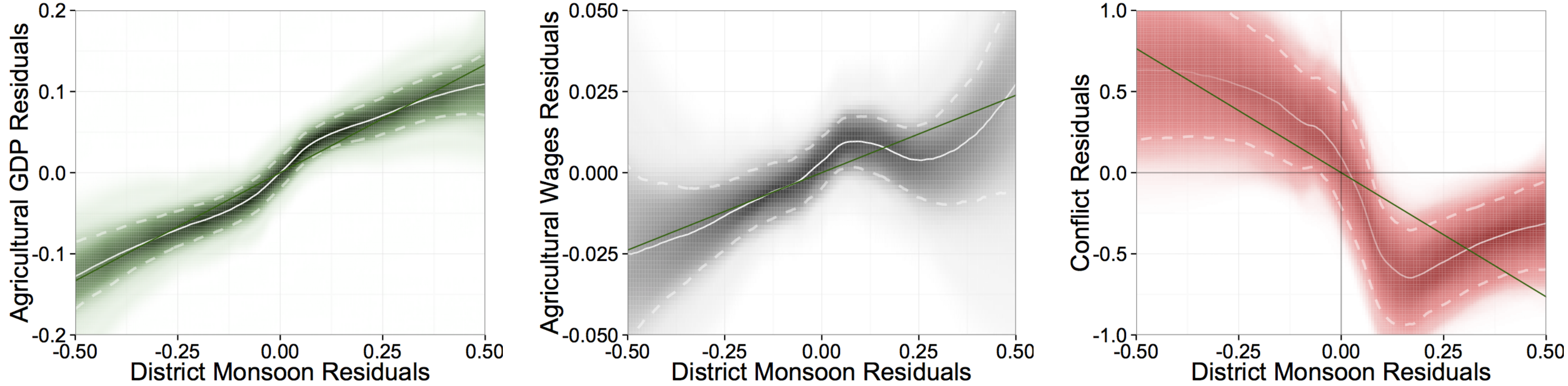
Panel B: Casualties



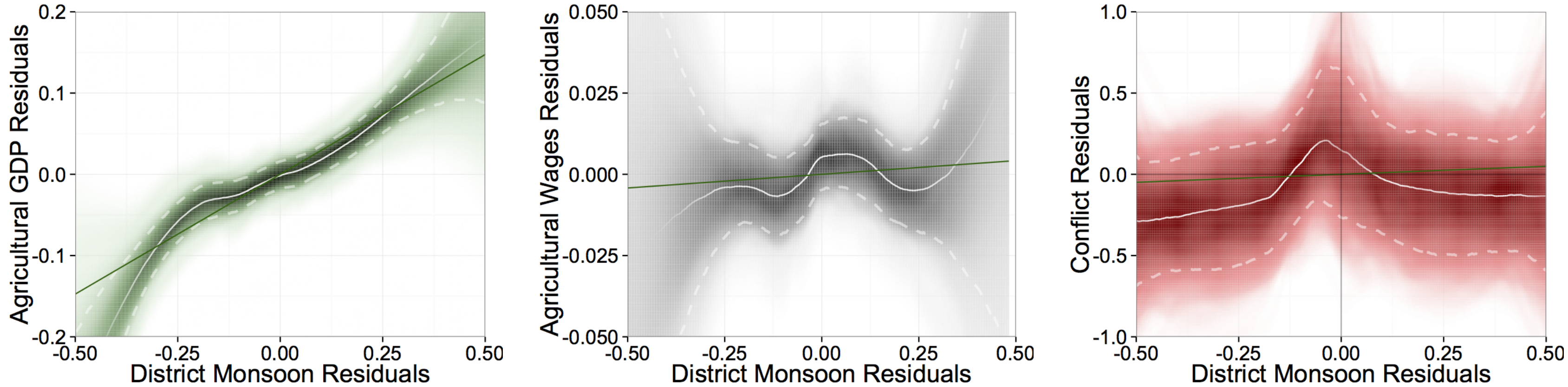
Notes: The vertical line indicates the NREGA introduction date. The blue solid lines indicate the coefficients for the pre- and post NREGA period obtained from a simple regression interacting lagged Monsoon rainfall with the NREGA treatment indicator. The red line are each point estimates of the relationship between lagged monsoon rainfall and conflict. 95% confidence bands are indicated as dotted black lines.

Figure 5: Non-parametric estimates of effect of monsoon rain on agricultural output per capita, wages and conflict before and after the introduction of NREGA.

Panel A: Before NREGA



Panel B: After NREGA

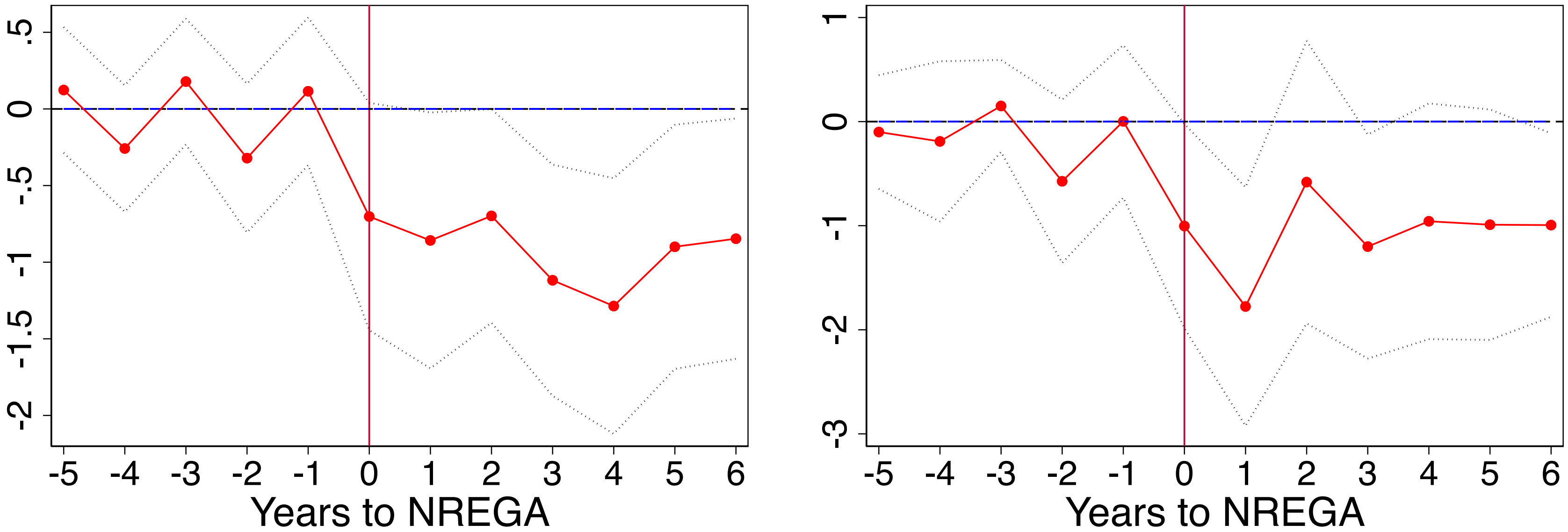


Notes: Figure plots non-parametric local linear regressions using the watercolor vizualization method as introduced in Hsiang et al. (2013): 95% confidence bands are indicated as dashed lines. Note that for both panels, the scales are identical, which allows a direct comparison. Data on the dependent variables (agricultural output per capita, wages and conflict events) as well as the independent variable monsoon rainfall has been demeaned, removing both district fixed effects as well as region by NREGA phase by time effects. The green line indicates the OLS fit, while the solid white line indicates the median bootstrapped non-linear curve.

Figure 6: Difference-in-Differences effect of the estimated NREGA monsoon rainfall elasticity on conflict levels relative to the NREGA introduction date

Panel A: Total Number of Events

Panel B: Casualties



Notes: The vertical line indicates the NREGA introduction date. The connected red dots are the point estimates of the interaction between the estimated district specific monsoon induced NREGA participation measure $\mathbb{1}(\widehat{\phi_d})$ estimated off the NREGA expenditure per district and year and the time to NREGA treatment indicator. 95% confidence bands obtained from clustering at the district level are indicated as dotted black lines.

Table 1: Comparing District Characteristics Across NREGA Implementation Phases

	Overall	Phase 1		Phase 2		Phase 3	
	Mean	Mean	p	Mean	p	Mean	p
Panel A: Conflict							
Red Corridor	0.379	0.559	0.000	0.375	0.911	0.228	0.000
	(0.486)	(0.498)	***	(0.486)		(0.421)	***
Conflict Events	1.477	1.894	0.128	1.645	0.679	1.033	0.040
	(4.542)	(5.044)		(5.232)		(3.578)	**
Casualties	2.702	3.660	0.079	2.719	0.983	1.874	0.042
	(8.786)	(10.384)	*	(10.358)		(5.867)	**
Panel B: Income							
Agricultural Output/Capita [INR]	320.722	219.651	0.000	268.988	0.163	433.085	0.000
	(453.169)	(292.799)	***	(419.184)		(548.632)	***
Agricultural Wage [INR]	69.932	52.586	0.000	67.559	0.521	84.593	0.000
	(36.712)	(15.773)	***	(34.650)		(42.859)	***
Lit Pixels [share]	0.522	0.383	0.000	0.455	0.008	0.675	0.000
	(0.335)	(0.300)	***	(0.302)	***	(0.319)	***
Panel C: Climate							
Monsoon Temperature	25.860	25.727	0.591	26.405	0.094	25.686	0.391
	(4.104)	(4.526)		(4.013)	*	(3.750)	
Annual Rain	1343.459	1357.093	0.710	1471.326	0.016	1264.499	0.025
	(677.649)	(586.596)		(648.074)	**	(753.074)	**
Monsoon Rain	961.201	1016.019	0.035	1048.984	0.020	868.116	0.000
	(481.958)	(412.287)	**	(465.144)	**	(529.685)	***
Drought Severity	-0.198	-0.215	0.241	-0.211	0.429	-0.177	0.068
	(0.231)	(0.262)		(0.195)		(0.220)	*
Panel D: Socio-Economic							
Illiterate [share]	0.465	0.523	0.000	0.471	0.572	0.413	0.000
	(0.120)	(0.106)	***	(0.129)		(0.103)	***
Household Size	5.432	5.390	0.371	5.519	0.194	5.421	0.802
	(0.827)	(0.801)		(0.846)		(0.838)	
Population younger than 6 [share]	0.250	0.261	0.000	0.253	0.282	0.240	0.000
	(0.034)	(0.026)	***	(0.033)		(0.037)	***
Population growth 1991-2001	22.087	21.599	0.373	24.463	0.021	21.253	0.116
	(10.540)	(8.095)		(13.551)	**	(10.445)	
Gender Gap [per 1000 people]	22.316	24.978	0.000	21.548	0.235	20.443	0.000
	(7.450)	(5.785)	***	(8.242)		(7.641)	***
Panel E: Infrastructure							
Primary School [share]	0.832	0.807	0.006	0.821	0.361	0.859	0.001
	(0.156)	(0.155)	***	(0.151)		(0.156)	***
Mudroad [share]	0.630	0.680	0.001	0.656	0.217	0.573	0.000
	(0.268)	(0.239)	***	(0.267)		(0.281)	***
Permanent Housing [share]	0.463	0.360	0.000	0.434	0.098	0.567	0.000
	(0.237)	(0.205)	***	(0.210)	*	(0.233)	***
Primary Health Care Centre [share]	0.391	0.321	0.000	0.373	0.320	0.461	0.000
	(0.250)	(0.220)	***	(0.228)		(0.266)	***
Electricity [share]	0.801	0.685	0.000	0.782	0.349	0.909	0.000
	(0.252)	(0.296)	***	(0.236)		(0.154)	***
Bus Stop [share]	0.444	0.334	0.000	0.397	0.052	0.564	0.000
	(0.320)	(0.286)	***	(0.293)	*	(0.320)	***
Postal Office [share]	0.490	0.373	0.000	0.465	0.230	0.603	0.000
	(0.274)	(0.236)	***	(0.252)		(0.272)	***
Panel F: NREGA							
Household Participating [share]	0.390	0.435	0.065	0.470	0.120	0.310	0.001
	(0.490)	(0.311)	*	(0.682)		(0.487)	***
Expenditure / Capita [INR]	541.045	595.253	0.196	678.879	0.121	422.020	0.006
	(848.051)	(540.597)		(1177.619)		(847.342)	***
Days Worked / Capita	3.494	4.016	0.048	4.142	0.212	2.704	0.003
	(5.197)	(3.668)	**	(6.786)		(5.281)	***

Notes: This table presents cross-sectional summary statistics at the district level. All time-varying variables are district level means, computed as an average for the period prior to NREGA, with the exception of the NREGA participation data in Panel F. Robust standard errors are in parentheses. The p-value are from a comparison of the respective covariate for district within an implementation phase, with all the other districts not in that implementation phase. Socio-economic and district Infrastructure statistics based on the 2001 Census for India. Infrastructure statistics is the share of villages in a district with access to a particular type of infrastructure.

Table 2: Reduced Form Relationship between Monsoon, Agricultural Production, Wages and Violence

	Agricultural Income		Conflict	
	(1)	(2)	(3)	(4)
	ln(Output/Capita)	ln(Wage)	Events	Casualties
$\log(\text{Monsoon}_t)$	0.325***	0.058**		
	(0.066)	(0.024)		
NREGA x $\log(\text{Monsoon}_t)$	-0.100	-0.085***		
	(0.078)	(0.027)		
$\log(\text{Monsoon}_{t-1})$			-1.402***	-1.406***
			(0.271)	(0.367)
NREGA x $\log(\text{Monsoon}_{t-1})$			1.298***	1.361***
			(0.323)	(0.401)
Joint Test:				
log(Monsoon)+ NREGA x log(Monsoon) = 0	.225***	-.027	-.103	-.045
	(.0864)	(.0298)	(.155)	(.323)
Observations	5507	3043	3522	3435
Number of Districts	471	371	239	239
Estimation	OLS	OLS	Poisson	Poisson

Notes: All regressions include region by NREGA phase and time effects and district fixed effects. Columns (1) and (2) study agricultural output and wages on an unbalanced annual district level panel from 1999-2011 and 1998-2010 respectively, using contemporaneous monsoon rainfall as independent variable. Columns (3) and (4) are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000- 2014. The dependent variables are number of conflict events and number of casualties. Note that district numbers for poisson model reflect the number of districts over the sample period for which there is variation in the dependent variable. For columns (1)-(2) standard errors are adjusted to reflect spatial dependence as modeled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. Poisson regressions present standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Decomposition of casualties and arrests: the impact of NREGA

	Baseline results		Casualty types			Arrests	No casualty events
	Events	Casualties	Civilians	Security forces	Militants		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Monsoon$_{t-1}$)	-1.402***	-1.406***	-1.210**	-1.800***	-1.820***	-1.142***	-1.167***
	(0.271)	(0.367)	(0.573)	(0.376)	(0.556)	(0.342)	(0.306)
NREGA x log(Monsoon$_{t-1}$)	1.298***	1.361***	0.827*	1.792***	2.214***	0.886**	1.207***
	(0.323)	(0.401)	(0.471)	(0.437)	(0.611)	(0.350)	(0.334)
Observations	3522	3435	2720	2503	2215	2961	3241
Number of Districts	239	239	202	176	159	209	221

Notes: All regressions include region-phase-time effects in addition to the district fixed effects. The different outcome measures are counts indicated in the column head, all regressions are estimated using a pseudo maximum likelihood poisson estimator. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Robustness to of NREGA Effect to different empirical models or the inclusion of other fixed effects or controls.

	(1)	(2)	(3)	(4)	(5)
Panel A: Events					
log(Monsoon$_{t-1}$)	-1.402***	-0.683***	-0.867***	-1.397***	-1.438***
	(0.271)	(0.227)	(0.214)	(0.254)	(0.288)
NREGA x log(Monsoon$_{t-1}$)	1.298***	0.534**	0.903***	1.293***	1.291***
	(0.323)	(0.260)	(0.245)	(0.306)	(0.340)
Joint Test:					
Pre NREGA + Post NREGA = 0	-.103	-.15	.037	-.104	-.147
	(.155)	(.171)	(.128)	(.152)	(.182)
Panel B: Casualties					
log(Monsoon$_{t-1}$)	-1.406***	-0.883*	-1.231***	-1.399***	-1.265***
	(0.367)	(0.478)	(0.359)	(0.348)	(0.358)
NREGA x log(Monsoon$_{t-1}$)	1.361***	0.624	1.314***	1.352***	1.285***
	(0.401)	(0.439)	(0.457)	(0.385)	(0.375)
Joint Test:					
Pre NREGA + Post NREGA = 0	-.045	-.259	.083	-.047	.02
	(.323)	(.348)	(.317)	(.317)	(.304)
District FE	X	X	X	X	X
Region x NREGA Phase x Year	X	X	X	X	.
Other controls	None	State x Year FE	District x NREGA	District Lin. Trend	Region x Year FE

Notes: All regressions include district fixed effects. Columns (1) - (4) all include region by NREGA phase by year time effects, while column (5) includes region by year FE and the NREGA treatment indicator. District X NREGA indicates a different set of district fixed effects for the period after NREGA was rolled out in a district. Standard errors are clustered at the district levels, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Robustness to controlling for state level police strength: Impact of NREGA on relationship between monsoon rainfall and conflict

	Events					Casualties				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log(Monsoon$_{t-1}$)	-0.985***	-1.031***	-0.963***	-0.978***	-1.013***	-1.364***	-1.393***	-1.366***	-1.255***	-1.270***
	(0.256)	(0.249)	(0.246)	(0.244)	(0.239)	(0.450)	(0.441)	(0.431)	(0.415)	(0.400)
NREGA x log(Monsoon$_{t-1}$)	1.029***	1.147***	1.129***	1.076***	1.089***	1.176***	1.307***	1.332***	1.193***	1.175***
	(0.298)	(0.283)	(0.267)	(0.258)	(0.255)	(0.418)	(0.416)	(0.407)	(0.406)	(0.403)
Civil Police strength	-0.044***	-0.043***	-0.041***	-0.041***	-0.041***	-0.051**	-0.051**	-0.050**	-0.053***	-0.054***
	(0.015)	(0.015)	(0.015)	(0.013)	(0.012)	(0.021)	(0.021)	(0.022)	(0.020)	(0.020)
NREGA x Civil Police strength	0.013**	0.013**	0.012**	0.012***	0.012***	0.015**	0.014**	0.014*	0.015**	0.016**
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
Armed Police strength		-0.004	-0.003	-0.005	-0.008		-0.029	-0.038	-0.036	-0.038
		(0.035)	(0.038)	(0.037)	(0.038)		(0.046)	(0.050)	(0.051)	(0.054)
NREGA x Armed Police strength		0.011	0.011	0.010	0.010		0.021	0.025	0.022	0.021
		(0.013)	(0.014)	(0.014)	(0.014)		(0.018)	(0.019)	(0.019)	(0.021)
Number of Police stations			0.015**	0.016***	0.015**			0.018**	0.020**	0.021**
			(0.006)	(0.006)	(0.006)			(0.008)	(0.008)	(0.009)
NREGA x Number of Police stations			-0.006**	-0.007**	-0.007**			-0.006	-0.009**	-0.010**
			(0.003)	(0.003)	(0.003)			(0.004)	(0.004)	(0.005)
Number of Police outposts				0.002	0.002*				0.001	0.001
				(0.001)	(0.001)				(0.002)	(0.002)
NREGA x Number of Police outposts				0.000	0.000				0.001	0.001
				(0.001)	(0.001)				(0.001)	(0.001)
Number of Armed batallions					0.053					-0.010
					(0.073)					(0.156)
NREGA x Number of Armed batallions					0.009					0.063
					(0.052)					(0.095)
Observations	2439	2440	2441	2441	2441	2226	2226	2226	2226	2226
Number of Districts	229	229	229	229	229	214	214	214	214	214

Notes: All regressions include region-phase-time effects and district fixed effects. The dependent variable throughout is the number of conflict events. Standard errors are clustered at the district level, with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Alternative Mechanism: Integrated Action Plan disbursals and the moderation of monsoon rainfall and conflict relationship

	IAP	Events			Casualties		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Monsoon$_{t-1}$)	0.234	-1.307***	-1.173***	-1.354***	-1.524***	-1.466***	-1.563***
	(0.214)	(0.254)	(0.207)	(0.278)	(0.374)	(0.354)	(0.336)
NREGA x log(Monsoon$_{t-1}$)		0.977***	1.136***	1.024***	0.776***	1.340***	1.389***
		(0.290)	(0.315)	(0.373)	(0.296)	(0.432)	(0.386)
IAP Expenditure				0.084***			0.055**
				(0.029)			(0.027)
Joint Test:							
log(Monsoon)+ NREGA x log(Monsoon) = 0		-.33*	-.037	-.33	-.748**	-.126	-.174
		(.182)	(.278)	(.224)	(.322)	(.384)	(.307)
Observations	259	3054	2493	2802	2951	2257	3121
Number of Districts	75	209	232	222	199	215	233

Notes: All regressions include region-phase-time effects and district fixed effects. Regressions (2) - (7) are estimated using a pseudo maximum likelihood poisson estimator. IAP expenditure data is available from 2010- 2013. Monsoon rain is the previous growing season's Monsoon rainfall realization. Column (1) studies IAP expenditure as a function of lagged Monsoon rain. Columns (2) and (5) remove the 33 districts that received the IAP originally. Columns (3) and (6) restrict the analysis to the period 2000-2010, before IAP started. Columns (4) and (7) controls for IAP expenditure. Standard errors are clustered at the district level, with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Ruling out alternative mechanisms

	Non-monsoon season	PMGSY Roads		Mining Sector		IAP Expenditure		All Combined	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(Outside Monsoon$_{t-1}$)	-0.152							0.271	0.455**
	(0.180)							(0.212)	(0.219)
NREGA x log(Outside Monsoon$_{t-1}$)	0.133							0.056	-0.215
	(0.199)							(0.240)	(0.248)
log(Monsoon$_{t-1}$)		-0.627***	-1.553***	-0.628***	-1.697***	-0.699***	-1.405***	-1.027***	-1.931***
		(0.143)	(0.226)	(0.128)	(0.254)	(0.130)	(0.226)	(0.141)	(0.204)
NREGA x log(Monsoon$_{t-1}$)			1.443***		1.504***		1.160***		1.539***
			(0.255)		(0.326)		(0.302)		(0.244)
Cumulative Roads		-11.201***	1.273					-13.177***	-0.625
		(3.606)	(2.939)					(3.674)	(2.972)
Cumulative Roads x log(Monsoon)		1.431***	-0.356					1.728***	-0.077
		(0.514)	(0.423)					(0.530)	(0.433)
Mining Sector x log(Monsoon)				1.429	0.793			1.269	1.525
				(1.944)	(1.697)			(2.260)	(1.872)
IAP Expenditure						-0.624**	-0.279	-0.247	0.036
						(0.297)	(0.301)	(0.415)	(0.393)
IAP Expenditure x log(Monsoon)						0.102**	0.051	0.045	0.003
						(0.042)	(0.043)	(0.060)	(0.057)

Notes: The dependent variable is the number of violent events reported on. All regressions include region by NREGA phase and time effects and district fixed effects and are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel. The number of districts and observation varies across specifications, as not for all mechanisms explored, data is available for the whole sample period from 2000 - 2014. Column (9) highlights that the results are robust to restricting the samples to the set of districts, for which data is available on all other mechanisms explored. Standard errors are clustered at the district level, with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Heterogeneity of the NREGA Effect by Implementation Phase and Inside the Red Corridor

	Conflict	
	(1) Events	(2) Casualties
Panel A: Overall NREGA Effect		
log(Monsoon$_{t-1}$)	-1.402***	-1.406***
	(0.271)	(0.367)
NREGA x log(Monsoon$_{t-1}$)	1.298***	1.361***
	(0.323)	(0.401)
Joint Test:		
	-.103	-.045
	(0.155)	(0.323)
Panel B: By NREGA Phase		
Phase 1 x log(Monsoon$_{t-1}$)	-2.097***	-1.896***
	(0.370)	(0.592)
NREGA x Phase 1 * log(Monsoon$_{t-1}$)	2.272***	2.412***
	(0.454)	(0.697)
Joint Test:	.176	.516
	(0.171)	(0.378)
Phase 2 x log(Monsoon$_{t-1}$)	-1.480**	-1.509**
	(0.660)	(0.689)
NREGA x Phase 2 * log(Monsoon$_{t-1}$)	1.833***	1.428
	(0.620)	(0.965)
Joint Test:	.353	-.081
	(0.259)	(0.524)
Phase 3 x log(Monsoon$_{t-1}$)	-0.790**	-1.298**
	(0.384)	(0.659)
NREGA x Phase 3 * log(Monsoon$_{t-1}$)	0.061	0.187
	(0.344)	(0.375)
Joint Test:	-.729***	-1.111*
	(0.263)	(0.601)
Panel C: In Red Corridor		
Inside Red Corridor x log(Monsoon$_{t-1}$)	-2.158***	-2.148***
	(0.264)	(0.522)
NREGA x Inside Red Corridor x log(Monsoon$_{t-1}$)	2.455***	2.943***
	(0.337)	(0.676)
Joint Test:	.297**	.794**
	(0.148)	(0.375)
Outside Red Corridor x log(Monsoon$_{t-1}$)	-0.989**	-1.409***
	(0.384)	(0.505)
NREGA x Outside Red Corridor x log(Monsoon$_{t-1}$)	0.363	0.309
	(0.321)	(0.354)
Joint Test:	-.626**	-1.1**
	(0.255)	(0.439)

Notes: All regressions include region by NREGA phase and time effects and district fixed effects. Regressions are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000- 2014. The dependent variables are number of conflict events and number of casualties (fatal and injured). *Joint Test* performs F-tests on the joint significance of the sum of the respective pairs of coefficients. Note that district numbers for poisson model reflect the number of districts over the sample period for which there dis variation in the dependent variable. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Explaining the NREGA Effect: Monsoon Rainfall and NREGA Participation

	(1) Expenditure	(2) Person Days	(3) Households
Panel A: Overall response			
log(Monsoon$_{t-1}$)	-108.171**	-0.844***	-0.023*
	(44.763)	(0.322)	(0.013)
Panel B: By NREGA Phase			
Phase 1 x log(Monsoon$_{t-1}$)	-138.605***	-0.811***	-0.046**
	(39.300)	(0.302)	(0.018)
Phase 2 x log(Monsoon$_{t-1}$)	-94.845	-1.258	0.001
	(116.411)	(0.812)	(0.019)
Phase 3 x log(Monsoon$_{t-1}$)	-82.362*	-0.640**	-0.014
	(43.186)	(0.256)	(0.019)
Panel C: In the Red Corridor			
Inside Red Corridor x log(Monsoon$_{t-1}$)	-127.226***	-1.033***	-0.032*
	(39.523)	(0.294)	(0.017)
Outside Red Corridor x log(Monsoon$_{t-1}$)	-94.549	-0.715*	-0.017
	(59.986)	(0.410)	(0.015)
Mean of DV	547	3.54	.395
Observations	3712	4242	4241

Notes: All regressions include region-phase-time effects and a set of district fixed effects for the period before 2010 and the period after 2010. Monsoon rain is the previous growing season's monsoon rainfall realization. Column (1) studies the total expenditure per capita, column (2) studies the number of days per capita in a district, while column (3) studies the share of a districts households that participate. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids. stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Insurance Value of NREGA: Monsoon Rainfall, Output Losses and NREGA Expenditures

	Output Value/Capita	NREGA Expenditure/Capita	
	(1)	(2)	(3)
	OLS	OLS	IV
log(Monsoon$_t$)	897.813***		
	(235.687)		
log(Monsoon$_{t-1}$)		-108.171**	
		(44.763)	
log(Agricultural Output Value/Capita$_{t-1}$)			-0.196***
			(0.044)
First Stage			39.4
Observations	5105	3712	1873
Number of Districts			425

Notes: All regressions include region-phase-time effects and district fixed effects. Column (1) relates monsoon rainfall with agricultural output per capita. Column (2) studies lagged monsoon rainfall and its effect on levels of NREGA expenditure in a district per capita. Column (3) is an instrumental variables exercise, instrumenting lagged agricultural output value per capita with lagged monsoon rainfall. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids. stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: Monsoon rainfall and incomes before and after NREGA introduction in rural areas: evidence from household panel data

	log(Household income in last year)					
	Whole population			Agricultural laborers		
	(1)	(2)	(3)	(4)	(5)	(6)
log(Monsoon$_{t-1}$)	0.157***	0.161***	0.187***	0.261***	0.273***	0.184*
	(0.041)	(0.041)	(0.042)	(0.072)	(0.088)	(0.099)
NREGA × log(Monsoon$_{t-1}$)	0.003	0.024	0.004	-0.123**	-0.112*	-0.167**
	(0.040)	(0.039)	(0.040)	(0.061)	(0.066)	(0.071)
Joint Test:						
log(Monsoon)+ NREGA x log(Monsoon) = 0	.16***	.185***	.191***	.138	.162	.016
	(.0579)	(.0596)	(.0607)	(.108)	(.121)	(.131)
Observations	50234	50231	45750	8884	8708	4420
Clusters	263	263	251	246	241	196
Fixed Effect	District	Village	Household	District	Village	Household

Notes: All regressions include region-phase-time effects in addition to the location or household fixed effects indicated in the column. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 12: Difference in Difference: NREGA Effect on conflict levels due to insurance

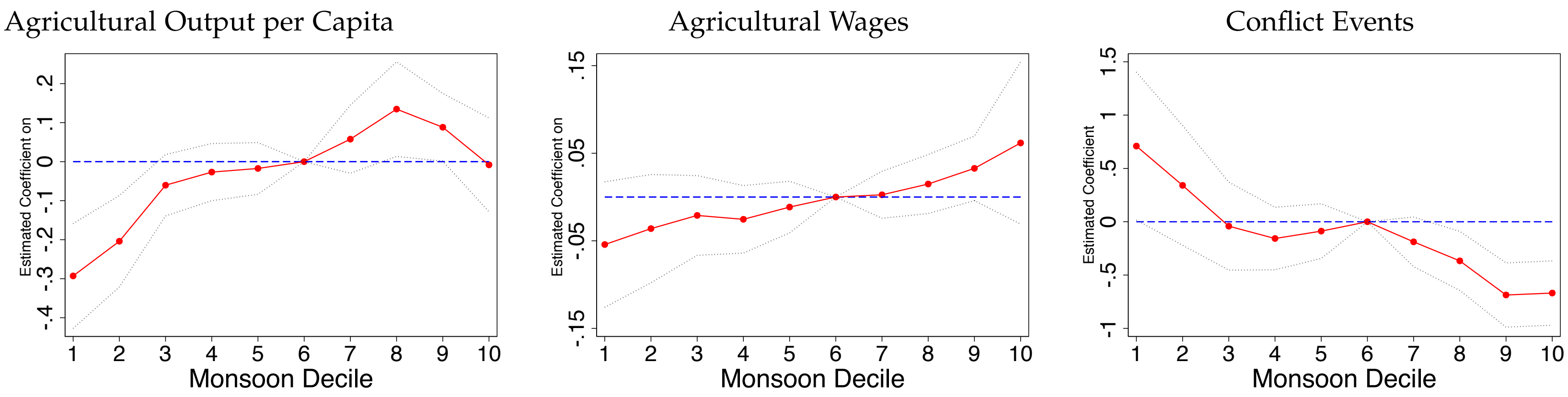
	Conflict Events			Casualties		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall Effect						
NREGA x $\mathbb{1}(\widehat{\phi_d})$	-0.925**	-0.629	-0.647**	-0.988**	-0.824*	-0.641
	(0.363)	(0.399)	(0.325)	(0.481)	(0.476)	(0.398)
Panel B: By NREGA Phase						
NREGA x Phase 1 x $\mathbb{1}(\widehat{\phi_d})$	-1.398***	-0.655	-0.872**	-1.191**	-0.749	-0.704
	(0.396)	(0.464)	(0.377)	(0.548)	(0.568)	(0.447)
NREGA x Phase 2 x $\mathbb{1}(\widehat{\phi_d})$	0.668	-0.533	1.128*	0.360	-0.099	1.942*
	(0.483)	(0.432)	(0.626)	(0.784)	(1.240)	(1.063)
NREGA x Phase 3 x $\mathbb{1}(\widehat{\phi_d})$	-0.189	-0.460	0.080	-0.599	-1.347**	-0.842
	(0.298)	(0.371)	(0.294)	(0.786)	(0.583)	(0.641)
Panel C: In the Red Corridor						
Inside Red Corridor x NREGA x $\mathbb{1}(\widehat{\phi_d})$	-1.308***	-0.793*	-1.073***	-1.315***	-1.047**	-1.065**
	(0.366)	(0.426)	(0.406)	(0.508)	(0.520)	(0.521)
Outside Red Corridor x NREGA x $\mathbb{1}(\widehat{\phi_d})$	0.286	0.606	0.359	0.057	0.201	0.147
	(0.334)	(0.529)	(0.235)	(0.471)	(0.710)	(0.314)
Mean of DV	4.55	4.55	4.55	6.88	6.88	6.88
Observations	3522	3522	3522	3435	3435	3435
Number of Districts	239	239	239	239	239	239
$\widehat{\phi_d}$ estimated off NREGA data for	Expenditure	Person days	Households	Expenditure	Person days	Households

Notes: Table presents results from a difference-in-difference exercise, comparing conflict levels in districts, across places where demand for NREGA is responsive to monsoon rainfall, before and after NREGA was introduced. $\mathbb{1}(\widehat{\phi_d}) = 1$ in case demand for NREGA is a decreasing function of monsoon rainfall. All regressions include region-phase-time effects and a set of district fixed effects, which ensures that the difference-in-difference is not identified off variation across NREGA implementation phases. Regressions are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000- 2014. The $\mathbb{1}(\widehat{\phi_d})$ is estimated using data on total expenditure per capita (columns (1) and (3)), the number of days worked per capita (columns (2) and (5)) and the share of households that demand NREGA work in a district and financial year (column (3) and (6)). Panel A presents the baseline results. Panel B explores heterogeneity by NREGA implementation phase, while Panel C explores the effect within the Red Corridor. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

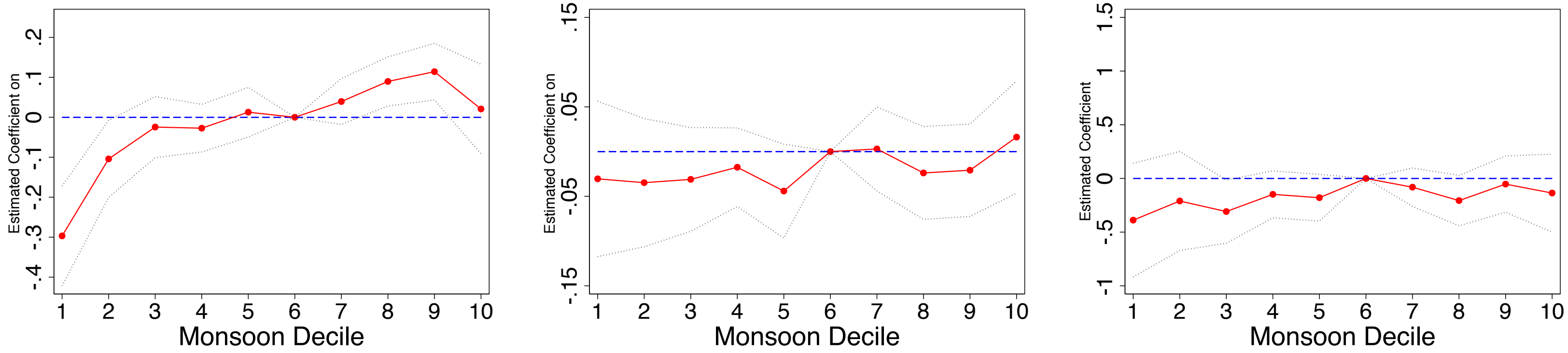
Figures and Tables for the Appendix

Figure A1: Non-parametric regressions of monsoon rainfall deciles on agricultural output per capita, agricultural wages (middle) and conflict events (right).

Panel A: Before NREGA



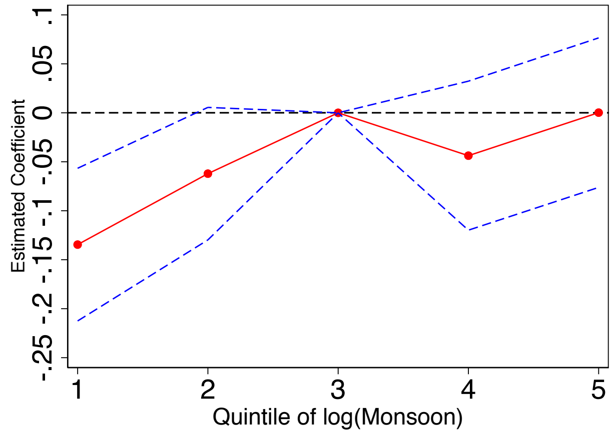
Panel B: After NREGA



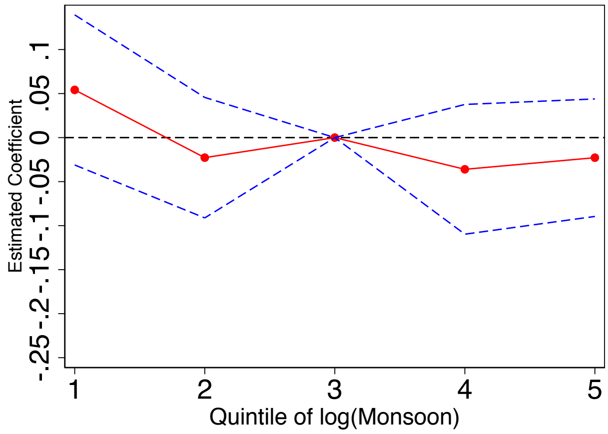
Notes: Non-parametric regressions including district and region-by-NREGA phase and time effects. The monsoon rainfall measure is broken out into deciles and the regressions are estimated for the two subsamples, before and after NREGA was introduced. Column (1) and (2) use the contemporaneous monsoon rainfall deciles, while column (3) uses the lagged monsoon rainfall deciles. The estimated coefficients are plotted with 90% confidence bands obtained from clustering standard errors at the district level indicated as dashed lines.

Figure A2: Effect of Monsoon Rain on agricultural wages before and after NREGA was introduced in the IDHS village panel

Panel A: before NREGA

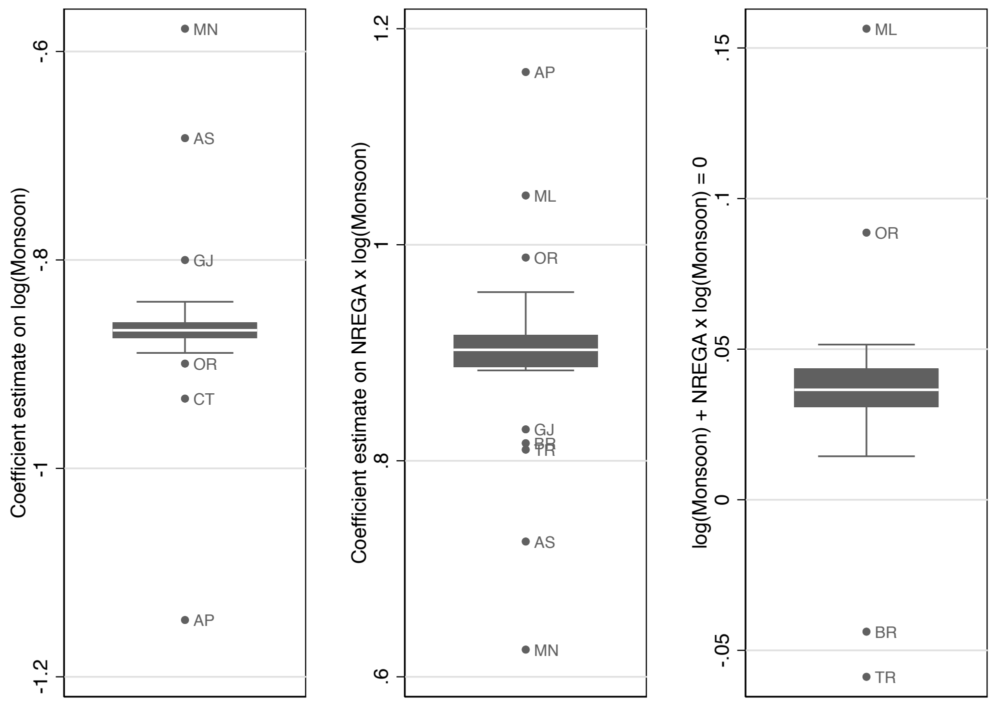


Panel B: after NREGA



Notes: The figure presents results from a regression including village and region by NREGA implementation phase by survey wave fixed effects. The graph plots out the point estimates of the effect of different quintiles of the monsoon rainfall distribution on agricultural wages. Standard errors are clustered at the district level with 90% confidence bands indicated.

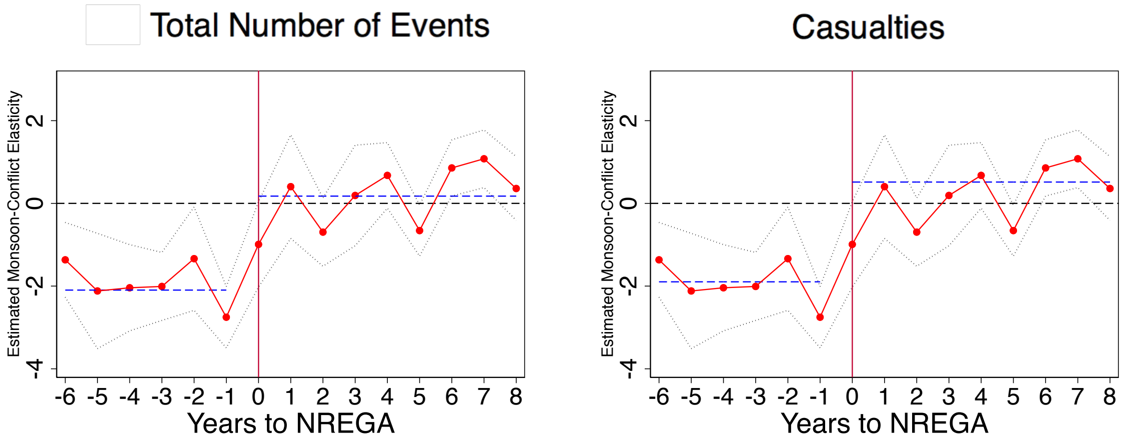
Figure A3: Robustness to dropping each state in turn



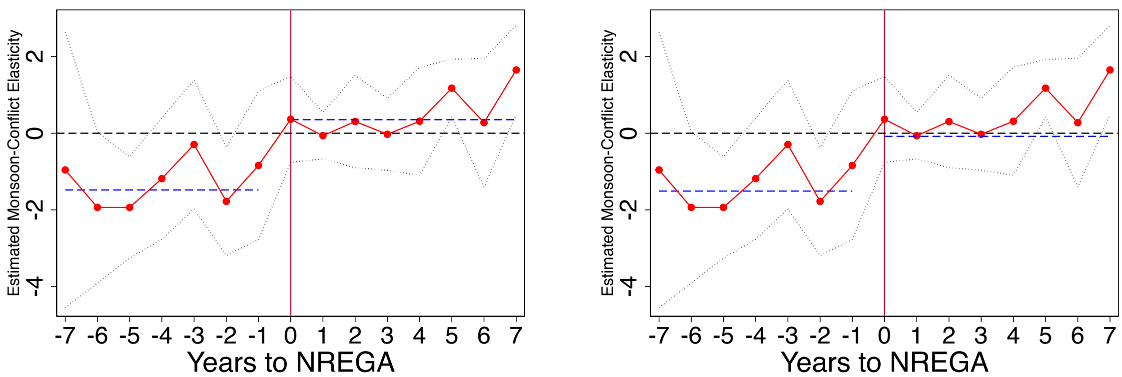
Notes: The figure presents the distribution of the estimated coefficients obtained when dropping each state in turn as a box plot. The first box plot presents the estimated coefficients on the log(monsoon rain) measure prior to NREGA introduction, the second presents the coefficient on the NREGA dummy interacted with the log(monsoon rain) measure, while the third figure presents the distribution of point estimates of the joint effect.

Figure A4: Effect of monsoon rain on conflict over time relative to NREGA introduction by NREGA implementation phase

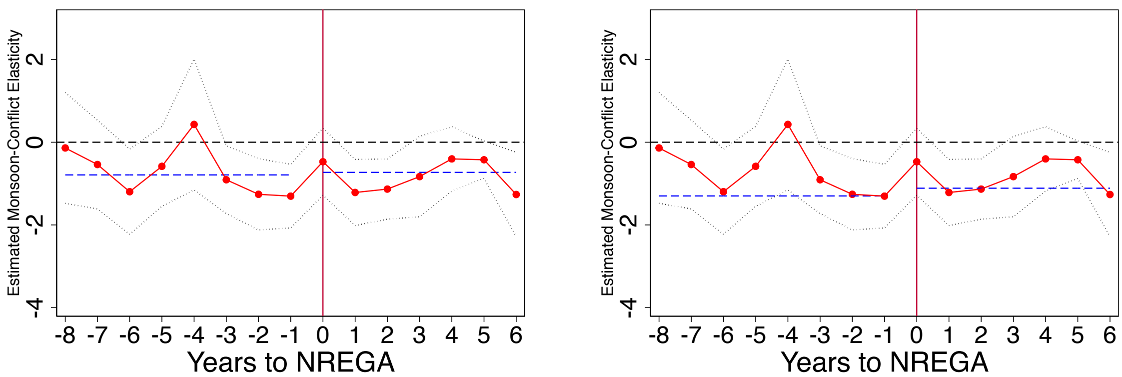
Panel A: NREGA Phase 1



Panel B: NREGA Phase 2



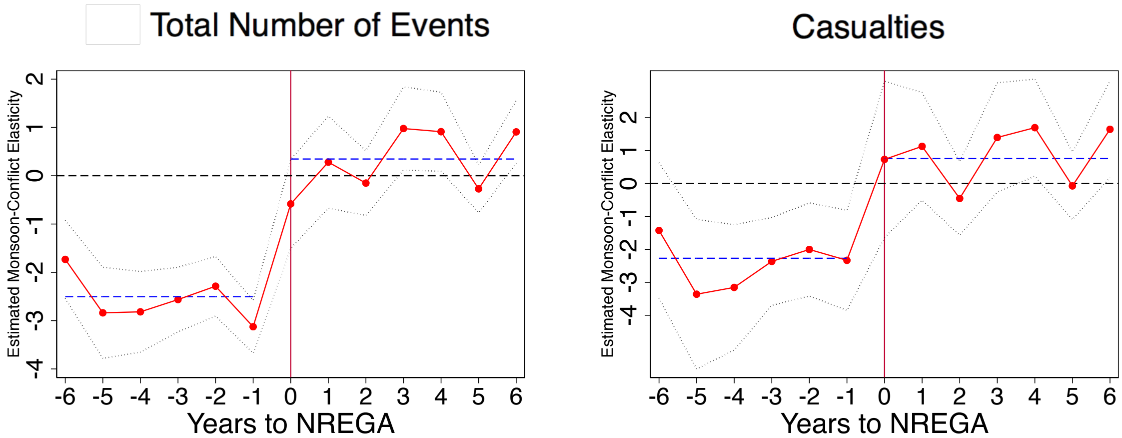
Panel C: NREGA Phase 3



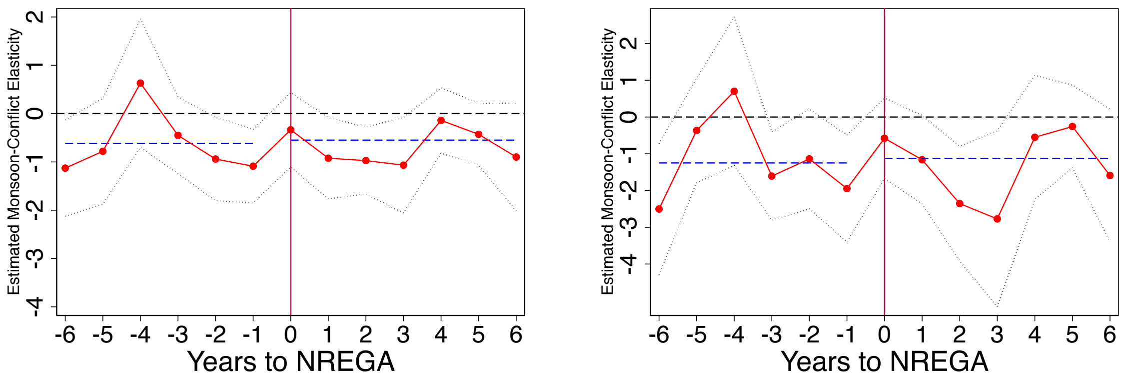
Notes: The vertical line indicates the NREGA introduction date. The vertical line indicates the NREGA introduction date. Each graph presents the result of a separate regression on the relevant sub-sample, controlling for district and region by NREGA implementation phase time effects. The connected red dots are the point estimates of the interaction between lagged monsoon rainfall with the time to NREGA treatment indicator. The dashed blue lines represent the pooled point estimates for the pre and post NREGA period, 95% confidence bands obtained from clustering at the district level are indicated as dotted black lines.

Figure A5: Effect of monsoon rain on conflict over time in the red corridor (panel A) compared to the rest of India (panel B).

Panel A: Red Corridor Districts

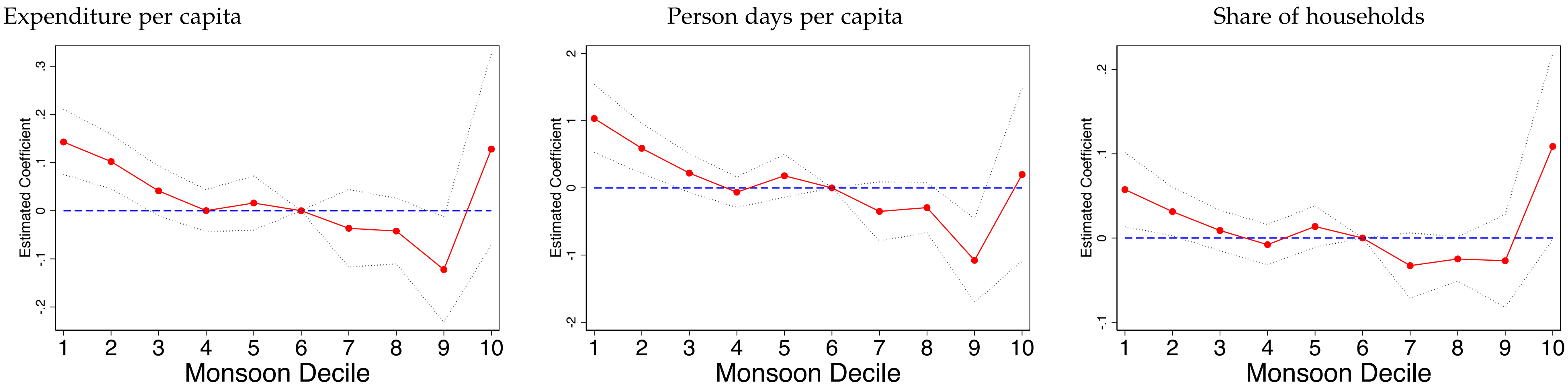


Panel B: Non-Red Corridor Districts



Notes: The vertical line indicates the NREGA introduction date. Each graph presents the result of a separate regression on the relevant sub-sample, controlling for district and region by NREGA implementation phase time effects. The connected red dots are the point estimates of the interaction between lagged monsoon rainfall with the time to NREGA treatment indicator. The dashed blue lines represent the pooled point estimates for the pre and post NREGA period, 95% confidence bands obtained from clustering at the district level are indicated as dotted black lines.

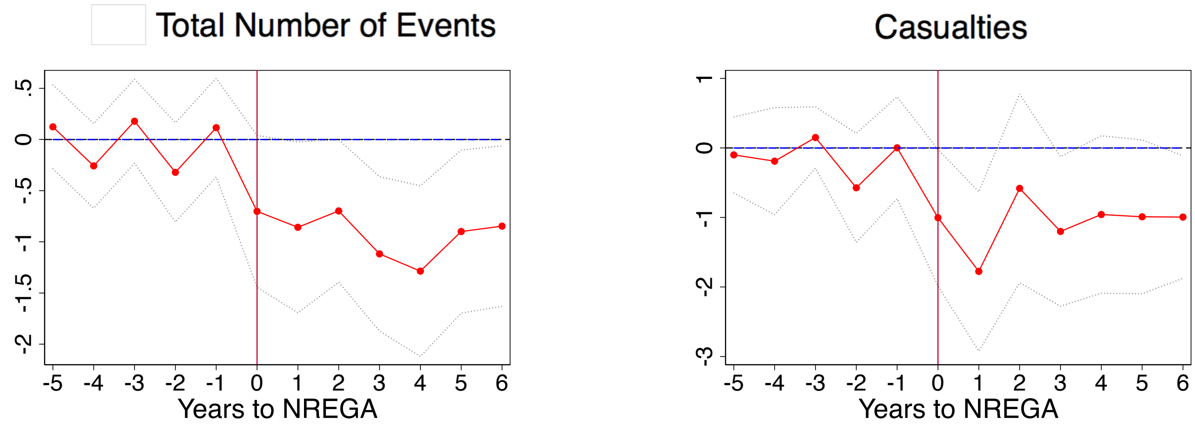
Figure A6: Non-parametric regressions of monsoon rainfall deciles on NREGA participation



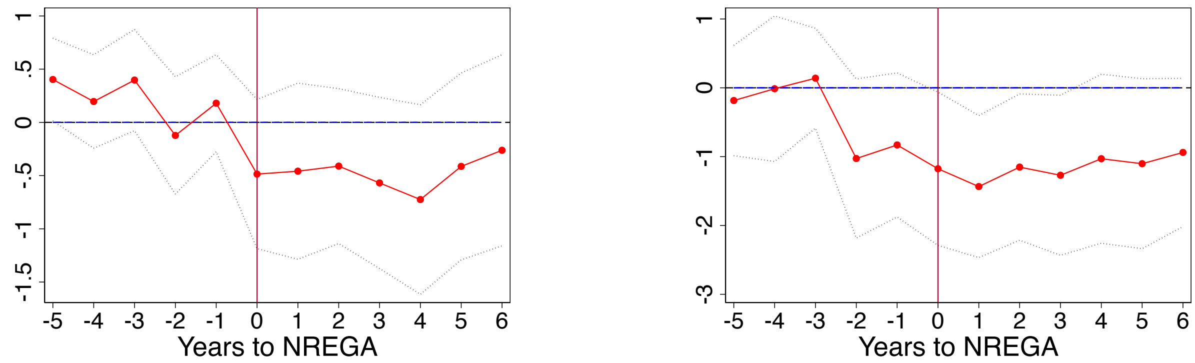
Notes: The figure presents results from a regression including district and region by NREGA implementation phase by year time effects. The graph plots out the point estimates of the effect of different deciles of the monsoon rainfall distribution on NREGA participation. The different NREGA participation measures are indicated in the figure head. Standard errors are clustered at the district level with 90% confidence bands indicated.

Figure A7: Effect of NREGA monsoon rainfall participation elasticity on conflict levels relative to the NREGA introduction date

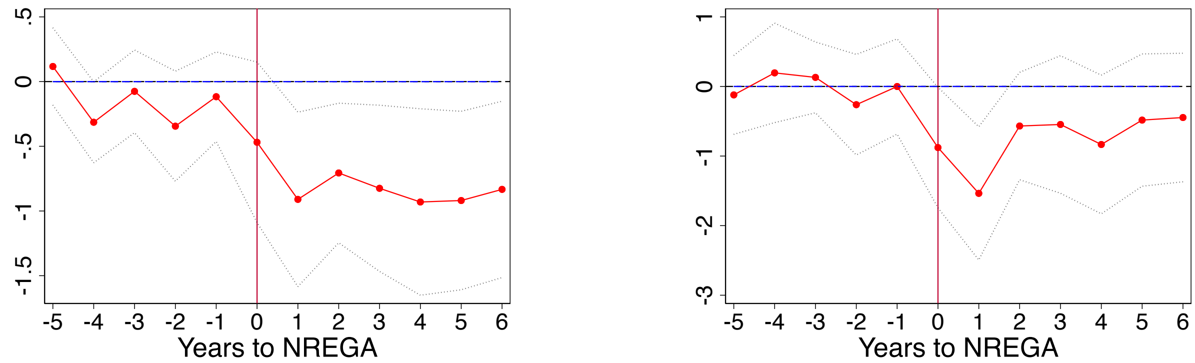
Panel A: $\mathbb{1}(\widehat{\phi_d})$ estimated off Expenditure/ Capita



Panel B: $\mathbb{1}(\widehat{\phi_d})$ estimated off Persondays/ Capita



Panel C: $\mathbb{1}(\widehat{\phi_d})$ estimated off Share of Households



Notes: Effect of NREGA on Conflict Levels: Figures present results from difference in difference estimation of the monsoon rainfall sensitivity of NREGA participation with regards to conflict levels relative to the introduction period indicated by the red vertical line. The connected red dots are the point estimates of the interaction between the estimated district specific monsoon induced NREGA participation measure $\mathbb{1}(\widehat{\phi_d})$ and the time to NREGA treatment indicator, 95% confidence bands obtained from clustering at the district level are indicated as dotted black lines.

Table A1: Monsoon rainfall and agricultural wages - village level survey data from the IDHS

	log(Agricultural worker wages)			log(Skilled construction workers)		
	(1)	(2)	(3)	(4)	(5)	(6)
log(Monsoon$_{t-1}$)	0.111***	0.110***	0.122***	0.191***	0.191***	0.200***
	(0.030)	(0.028)	(0.029)	(0.037)	(0.035)	(0.031)
NREGA $\times$ log(Monsoon$_{t-1}$)	-0.113***	-0.099***	-0.115***	-0.051	-0.047	-0.060
	(0.028)	(0.029)	(0.032)	(0.038)	(0.038)	(0.038)
Joint Test:						
log(Monsoon)+ NREGA x log(Monsoon) = 0	-.002	.011	.008	.14**	.144***	.14***
	(.0392)	(.0388)	(.0433)	(.0563)	(.0548)	(.051)
Observations	2617	2564	2275	2728	2679	2430
Clusters	250	250	221	255	255	227
Location FE	District	Village	Village	District	Village	Village
Village controls	No	No	Yes	No	No	Yes

Notes: All regressions include region-phase-time time effects in addition to the location or household fixed effects indicated in the column. Village controls control for a set of time varying village level measures capturing public good access, such as distance to nearest police station, infrastructure, whether villages have road access, electricity or have self-help groups. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Robustness to using the Gawande et al. (2017) alternative conflict data leveraging local languages

	(1) Overall	(2) Civilian	(3) Maoist	(4) Security forces
log(Monsoon$_{t-1}$)	-1.304***	-2.223***	-0.815	-1.343
	(0.489)	(0.559)	(0.672)	(0.852)
NREGA x log(Monsoon$_{t-1}$)	2.952***	4.820***	1.546	3.182**
	(1.019)	(1.210)	(1.044)	(1.586)
Observations	504	465	419	428
Number of Districts	56	52	47	48
Estimation	Poisson	Poisson	Poisson	Poisson

Notes: All regressions include region-phase-time time effects and a set of district fixed effects. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Robustness to Alternative Weather Measures: Moderating Effect of NREGA on Conflict

	Events					Casualties				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Monsoon deficiency$_{t-1}$	1.798***					1.768***				
	(0.287)					(0.499)				
NREGA x Monsoon deficiency$_{t-1}$	-1.686***					-1.672***				
	(0.283)					(0.573)				
Normalized Monsoon$_{t-1}$		-0.148***					-0.107			
		(0.050)					(0.070)			
NREGA x Normalized Monsoon$_{t-1}$		0.142**					0.086			
		(0.055)					(0.074)			
Negative Monsoon$_{t-1}$			0.319***					0.210		
			(0.101)					(0.203)		
NREGA x Negative Monsoon$_{t-1}$			-0.489***					-0.639**		
			(0.137)					(0.271)		
Positive Monsoon$_{t-1}$			-0.415***					-0.299		
			(0.139)					(0.283)		
NREGA x Positive Monsoon$_{t-1}$			0.309**					0.066		
			(0.148)					(0.276)		
log(GPCC Rain$_{t-1}$)				-1.193***					-1.383***	
				(0.238)					(0.333)	
NREGA x log(GPCC Rain$_{t-1}$)				1.488***					1.582***	
				(0.291)					(0.387)	
Drought Severity$_{t-1}$					6.678***					7.330***
					(1.661)					(2.197)
NREGA x Drought Severity$_{t-1}$					-9.314***					-10.330***
					(2.730)					(3.807)
Observations	3522	3522	3522	3270	2761	3435	3435	3435	3121	2602
Number of Districts	239	239	239	238	234	239	239	239	233	226

Notes: All regressions include region by NREGA phase and time time effects and district fixed effects. Regressions are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000-2014. The number of districts reported corresponds to the districts for which the dependent variable, conditional on the fixed effects, has any variation over the sample period. The weather measures are lagged by one year. Columns (1) and (7) present results where Monsoon rainfall is normalized by its sample standard deviation. Columns (2) and (8) are placebo's studying rainfall outside the monsoon growing season. Columns (3) and (9) use the GPCC rainfall data as alternative rainfall data source, which is available from 2000-2010. Columns (4) and (10) instrument the TRMM rainfall data with the GPCC data to remove measurement error. Columns (5) and (11) use the Drought Severity index instrumented with monsoon rainfall, which is available for 2000-2011. Columns (6) and (12) use average temperatures during the growing season. Standard errors are clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Robustness to controlling for other fixed district characteristic interacted with the NREGA treatment indicator

	Events			Casualties		
	(1)	(2)	(3)	(4)	(5)	(6)
log(Monsoon$_{t-1}$)	-0.729***	-0.727***	-0.632***	-1.016***	-1.011***	-0.654
	(0.201)	(0.203)	(0.198)	(0.334)	(0.326)	(0.431)
NREGA x log(Monsoon$_{t-1}$)	0.720***	0.714***	0.742***	0.916***	0.905***	0.577
	(0.222)	(0.222)	(0.219)	(0.296)	(0.294)	(0.405)
Observations	3522	3522	2442	3435	3435	2225
Number of Districts	239	239	229	239	239	214
District Characteristics x NREGA	Yes	Yes	Yes	Yes	Yes	Yes
Star State x NREGA	No	Yes	Yes	No	Yes	Yes
Police Force & Deployment proxies x NREGA	No	No	Yes	No	No	Yes

Notes: All regressions include region-phase-time time effects and district fixed effects. The dependent variable throughout is the number of conflict events. Standard errors are clustered at the district level, with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Alternative Mechanism: Rural Connectivity and Moderation of Rainfall and Conflict Relationship

	Total Events		Casualties	
	(1)	(2)	(3)	(4)
Panel A: Road Construction				
log(Monsoon$_{t-1}$)	-0.624***	-1.518***	-0.905***	-1.683***
	(0.136)	(0.273)	(0.341)	(0.395)
NREGA x log(Monsoon$_{t-1}$)		1.304***		1.336***
		(0.322)		(0.426)
Roads	0.850***	1.292	0.720	-4.234
	(0.308)	(3.723)	(0.547)	(6.544)
Roads x log(Monsoon_$t-1$)		-0.072		0.708
		(0.526)		(0.923)
Joint Test:				
log(Monsoon)+ NREGA x log(Monsoon) = 0		-.213		-.347
		(.151)		(.343)
Observations	2485	2485	2289	2289
Number of Districts	230	230	217	217
Panel B: Cumulative Road Construction				
log(Monsoon$_{t-1}$)	-0.581***	-1.550***	-0.900***	-1.912***
	(0.133)	(0.226)	(0.334)	(0.406)
NREGA x log(Monsoon$_{t-1}$)		1.443***		1.694***
		(0.255)		(0.449)
Cumulative Roads	-1.123***	1.273	-0.815	7.235
	(0.434)	(2.937)	(0.609)	(7.549)
Cumulative Roads x log(Monsoon_$t-1$)		-0.356		-1.175
		(0.423)		(1.051)
Joint Test:				
log(Monsoon)+ NREGA x log(Monsoon) = 0		-.107		-.217
		(.154)		(.335)
Observations	2485	2485	2289	2289
Number of Districts	230	230	217	217

Notes: All regressions include region-phase-time effects and district fixed effects. Regressions are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000-2014. Monsoon rain is the previous growing season's Monsoon rainfall realization. Panel A studies the effect of contemporaneous road construction on violence, while Panel B studies the impact of rainfall through the overall share of unconnected villages, that became connected up to 2012. Standard errors are clustered at district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: Alternative Mechanism: Mining Sector Share, Commodity Boom and Moderation of Rainfall and Conflict Relationship

	Total Events		Casualties	
	(1)	(2)	(3)	(4)
$\log(\text{Monsoon}_{t-1})$	-0.628***	-1.697***	-0.505*	-1.516***
	(0.128)	(0.254)	(0.281)	(0.414)
NREGA x $\log(\text{Monsoon}_{t-1})$		1.504***		1.544***
		(0.326)		(0.418)
Mining Sector	-37.215**	-31.587**	-81.684***	-69.701***
	(14.906)	(12.973)	(18.281)	(17.240)
Mining Sector x log(Monsoon_t-1)	1.429	0.793	6.423***	5.343**
	(1.944)	(1.697)	(2.312)	(2.189)
Joint Test:				
log(Monsoon)+ NREGA x log(Monsoon) = 0		-.193		.029
		(.145)		(.284)
Observations	3259	3259	3180	3180
Number of Districts	221	221	221	221

Notes: All regressions include region-phase-time effects and district fixed effects. Regressions are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000-2014. Monsoon rain is the previous growing season's Monsoon rainfall realization. Mining Sector Share is the share of the districts domestic product that is generated in the Mining sector based on data between 1998 and 2005. Standard errors are clustered at the district level, with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Is NREGA demand led? Is rationing correlated with monsoon rainfall? Evidence from IDHS 2010/2011 data

	(1)	(2)	(3)
Panel A: log(NREGA income)			
$\log(\text{Monsoon}_{t-1})$	-0.140**	-0.122**	-0.704**
	(0.067)	(0.060)	(0.298)
Observations	8263	8263	8257
Clusters	218	218	212
Panel B: Rationing (Days eligible minus days actually worked)			
$\log(\text{Monsoon}_{t-1})$	7.870***	4.117**	16.973
	(2.460)	(2.063)	(10.703)
Observations	9535	9535	9529
Clusters	232	232	226
Panel C: Not enough work available			
$\log(\text{Monsoon}_{t-1})$	0.091*	0.106**	0.064
	(0.047)	(0.047)	(0.117)
Observations	8233	8233	8225
Clusters	229	229	221
Location FE	None	State	District

Notes: Data is an individual job card specific cross section from the second IDHS survey wave. All regressions include region-phase time effects and a set of district fixed effects. The dependent variable in Panel C is a dummy indicating whether individuals state that there is not sufficient work available. This is estimated on the subset of individuals who report that they have been awarded fewer days of work then they are eligible. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: In which districts is demand for NREGA work correlated with monsoon rainfall?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Household Size	-0.106***											
	(0.018)											
Illiterate [share]		0.154***	0.126***	0.105***	0.087***	0.090***	0.089***	0.083***	0.087***	0.167***	0.170***	0.163***
		(0.026)	(0.028)	(0.029)	(0.027)	(0.027)	(0.031)	(0.030)	(0.029)	(0.041)	(0.042)	(0.041)
Population younger than 6 [share]		-0.180***	-0.191***	-0.172***	-0.192***	-0.197***	-0.166***	-0.177***	-0.182***	-0.266***	-0.268***	-0.227***
		(0.029)	(0.029)	(0.030)	(0.029)	(0.029)	(0.035)	(0.033)	(0.033)	(0.046)	(0.047)	(0.049)
Gender Gap [per 1000 people]			0.063***	0.066***	0.106***	0.112***	0.079***	0.107***	0.115***	0.109***	0.108***	0.107***
			(0.021)	(0.021)	(0.022)	(0.023)	(0.023)	(0.025)	(0.026)	(0.033)	(0.033)	(0.033)
Inside Red Corridor				0.121***	0.135***	0.135***	0.134***	0.136***	0.139***	0.153***	0.157***	0.140***
				(0.037)	(0.036)	(0.036)	(0.040)	(0.040)	(0.040)	(0.051)	(0.051)	(0.051)
Tribal Population [share]					0.084***	0.086***		0.073***	0.075***	0.086*	0.081*	0.078*
					(0.018)	(0.019)		(0.020)	(0.020)	(0.045)	(0.045)	(0.043)
Elevation						-0.020			-0.024	-0.048**	-0.048**	-0.051**
						(0.015)			(0.015)	(0.022)	(0.022)	(0.022)
Primary School [share]							-0.064**	-0.079***	-0.073***	-0.083**	-0.080**	-0.081**
							(0.026)	(0.025)	(0.026)	(0.038)	(0.038)	(0.038)
Primary Health Care Centre [share]							0.063**	0.081***	0.080***	0.080**	0.075**	0.075**
							(0.025)	(0.025)	(0.025)	(0.035)	(0.035)	(0.035)
Monsoon elasticity of agricultural output							0.048***	0.043***	0.043***	0.071**	0.070**	0.064**
							(0.016)	(0.015)	(0.015)	(0.030)	(0.028)	(0.029)
Monsoon elasticity of agricultural wages										0.041**	0.044***	0.043***
										(0.016)	(0.015)	(0.015)
Phase 2											-0.097*	-0.095*
											(0.052)	(0.052)
Population growth 1991-2001												-0.072**
												(0.029)
Best Subset												X
Mean of DV	.199	.199	.199	.199	.199	.199	.191	.191	.191	.225	.225	.225
Observations	543	543	543	543	543	543	471	471	471	284	284	284
R2	.0707	.0869	.104	.124	.16	.163	.186	.21	.213	.302	.311	.322

Notes: Table reports results from OLS regressions. The dependent variable is the $\mathbb{1}(\widehat{\phi_d})$ dummy as described in Section 5, where the $\widehat{\phi_d}$ are estimated using the NREGA participation data measuring the number of days worked per capita. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. The full set of variables among which best-subset selection choses includes 31 covariates. The best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9: Difference in Difference: NREGA Effect on conflict levels due to insurance using alternative $\mathbb{1}(\widehat{\phi_d})$ definition.

	Conflict Events			Casualties		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: $\widehat{\phi_d}$ in the lower 10%						
NREGA x $\mathbb{1}(\widehat{\phi_d})$	-0.872**	-1.016**	-0.151	-0.821	-1.185**	-0.039
	(0.423)	(0.473)	(0.465)	(0.547)	(0.600)	(0.390)
Panel B: $\widehat{\phi_d}$ in the lower 20%						
NREGA x $\mathbb{1}(\widehat{\phi_d})$	-0.925**	-0.629	-0.647**	-0.988**	-0.824*	-0.641
	(0.363)	(0.399)	(0.325)	(0.481)	(0.476)	(0.398)
Panel C: $\widehat{\phi_d}$ in the lower 30%						
NREGA x $\mathbb{1}(\widehat{\phi_d})$	-0.627*	-0.469	-0.450*	-0.926**	-0.761**	-0.531
	(0.334)	(0.304)	(0.261)	(0.420)	(0.377)	(0.357)
$\mathbb{1}(\widehat{\phi_d})$ estimated off NREGA data for	Expenditure	Person days	Households	Expenditure	Person days	Households

Notes: Table presents results from a difference-in-difference exercise, comparing conflict levels in districts, across places where demand for NREGA is responsive to monsoon rainfall, before and after NREGA was introduced. $\mathbb{1}(\widehat{\phi_d})$ is a dummy variable that is coded as 1 in case the estimate $\widehat{\phi_d}$ is in the lower 20% of the empirical distribution of estimates of ϕ_d indicating that demand for NREGA is decreasing in monsoon rainfall in a district. All regressions include region-phase-time time effects and a set of district fixed effects, which ensures that the difference-in-difference is not identified off variation across NREGA implementation phases. Regressions are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000- 2014. The $\widehat{\phi_d}$ is estimated using data on total expenditure per capita (columns (1) and (3)), the number of days worked per capita (columns (2) and (5)) and the share of households that demand NREGA work in a district and financial year (column (3) and (6)). Panel A presents the baseline results. Panel B explores heterogeneity by NREGA implementation phase, while Panel C explores the effect within the Red Corridor. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A10: Difference in Difference: NREGA effect on conflict levels due to insurance using continuous ϕ_d measure

	Conflict Events			Casualties		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall Effect						
NREGA x $\widehat{\phi_d}$	-0.602	-0.086**	-0.827**	-0.945	-0.133**	-1.083
	(0.369)	(0.043)	(0.399)	(0.617)	(0.066)	(0.667)
Panel B: By NREGA Phase						
NREGA x Phase 1 x $\widehat{\phi_d}$	-1.459***	-0.113*	-1.272**	-1.439*	-0.145*	-1.205
	(0.519)	(0.058)	(0.631)	(0.791)	(0.077)	(0.860)
NREGA x Phase 2 x $\widehat{\phi_d}$	0.248	-0.041*	-0.531	0.111	-0.063*	-0.358
	(0.204)	(0.025)	(0.759)	(0.221)	(0.032)	(0.989)
NREGA x Phase 3 x $\widehat{\phi_d}$	-0.010	-0.017	0.057	-0.146	-0.081	-0.702
	(0.032)	(0.016)	(0.162)	(0.220)	(0.086)	(0.536)
Panel C: In the Red Corridor						
Inside Red Corridor x NREGA x $\widehat{\phi_d}$	-1.411***	-0.114**	-1.429**	-1.631**	-0.174**	-1.424
	(0.501)	(0.055)	(0.711)	(0.826)	(0.084)	(1.041)
Outside Red Corridor x NREGA x $\widehat{\phi_d}$	0.046	-0.011	0.001	-0.015	-0.018	-0.369
	(0.038)	(0.010)	(0.148)	(0.075)	(0.023)	(0.334)
Mean of DV	3.31	2.97	3.35	5.14	4.74	5.06
Observations	3492	3492	3492	3405	3405	3405
Number of Districts	237	237	237	237	237	237
$\mathbb{1}(\widehat{\phi_d})$ estimated off NREGA data for	Expenditure	Person days	Households	Expenditure	Person days	Households

Notes: Table presents results from a difference-in-difference exercise, comparing conflict levels in districts, across places where demand for NREGA is responsive to monsoon rainfall, before and after NREGA was introduced. $\widehat{\phi_d}$ measures whether demand for NREGA is decreasing in monsoon rainfall. Regressions are weighted by the inverse of the standard error of the estimated $\widehat{\phi_d}$. All regressions include region-phase-time time effects and a set of district fixed effects. Regressions are estimated using a pseudo maximum likelihood poisson estimator, on a balanced district level annual panel from 2000- 2014. The $\widehat{\phi_d}$ is estimated using data on total expenditure per capita (columns (1) and (3)), the number of days worked per capita (columns (2) and (5)) and the share of households that demand NREGA work in a district and financial year (column (3) and (6)). Panel A presents the baseline results. Panel B explores heterogeneity by NREGA implementation phase, while Panel C explores the effect within the Red Corridor. Standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix to "Can Workfare Programs Moderate Violence? Evidence from India"

For Online Publication

Thiemo Fetzer

August 2, 2019

A Appendix Extended Results and Further Robustness

A.1 Ruling Out Other Mechanisms

This section rules out a range of alternative explanations or policies that could explain why the Monsoon rainfall and conflict relationship has become weaker. Most notably is the Pradhan Mantri Gram Sadak Yojana Rural Road Construction Scheme (PMGSY) that was implemented around the same time as NREGA was devised and introduced. Other confounders are the Integrated Action Plan, which channels additional funds into left-wing extremist affected districts. I also address a concern that large mineral sectors are driving the observed moderation.

A.1.1 Pradhan Mantri Gram Sadak Yojana Rural Road Construction Scheme

A concern with the analysis is that the Indian government has put forth many other development programs, whose implementation may affect the relationship between Monsoon and conflict at the same time and may be correlated with the roll-out of NREGA. In this case, the results would falsely attribute the observed inward rotation of the Monsoon-rainfall andconflict relationship to the employment guarantee scheme. The most prominent developmental scheme that was implemented around the same time is the Pradhan Mantri Gram Sadak Yojana (PMGSY). This scheme was introduced in 2000 and aims to provide improved road access for rural households. The scheme in particular aimed to provide roads to all villages with at least 1000 inhabitants by 2003, with a population of 500 and more by 2007 and had special provisions for tiny villages with at least 250 inhabitants for the hill states, tribal areas and desert areas. These were to be connected by 2007. As early NREGA districts are among the poorest and least urbanised, they are more likely to have received treatment through the PMGSY as well, which could partly explain my reduced form findings.

The crucial role that transport infrastructure may have in mitigating adverse weather shocks has been highlighted in Burgess and Donaldson (2010). Aggarwal (2014) evaluates the impact of the PMGSY using a difference-in-difference design and finds that the scheme increased incomes by increasing the potential market size for locally produced agricultural commodities; in addition, there is less price dispersion across market centers. I use her data to see whether the PMGSY moderates the relationship between Monsoon rainfall and conflict. I construct two variables: first, the share of all unconnected habitats connected in a year and second, the cumulative share of habitats among the unconnected habitats that received road access by the end of each year. The former measure may pick up direct effects from road construction on violence, while the latter variable, in its interaction with rainfall, could pick up the more persistent effects of this scheme by connecting previously unconnected villages.

The empirical design is identical to the main analysis, except that I now add these controls and interaction terms to the main specification. The results are presented in Table A5. Column (1) and (2) study violence intensity, while column (3) and (4) look at incidence. Panel A presents the results for contemporary road construction, while Panel B looks at cumulative connectivity. Columns (1) and (3) look at the rural connectivity and its interaction with rainfall by itself, while column (3) and (4) are a type of horse race. In neither specifications do the road construction interactions with rainfall achieve predictive power. This renders me confident that my results genuinely reflect the effect of the workfare scheme on the dynamics of conflict.

A.1.2 Mineral resources

Another concern is that the NREGA interactions may be picking up moderation of rainfall shocks due to a sectoral shift away from agriculture to the mining sector, which is less affected by rainfall variation. Vanden Eynde (2016) shows that districts with a large mining sector see a smaller elasticity between rainfall and conflict. If the introduction of NREGA is correlated with a sectoral shift towards the mineral resource sector, the NREGA interactions could be picking up this effect. This is not entirely implausible as the mid 2000s saw a commodity price boom which could have induced a lot more investment in the mining sector. In order to control for this I construct a share of a district's income that is due to the mining sector.[1]

Again, the specifications I present are very similar, adding a simple interaction with the mining sector share in district domestic product interacted with the Monsoon season rainfall. The results are presented in Table A6. Column (1) presents the results on violence intensity without the NREGA interactions. It becomes evident that

[1] I use district domestic product data for years between 1998 to 2005 available from `http://planningcommission.nic.in/plans/stateplan/index.php?state=ssphdbody.htm`, accessed on 21.06.2014. The district domestic product construction is discussed in detail in Katyal et al. (2001).

districts with a larger share of the mining sector experience a weaker relationship between violence and Monsoon rain. This maps into the findings of Vanden Eynde (2016). Once including the NREGA interaction, the coefficient on the Mining sector interaction becomes insignificant. More importantly, the NREGA interaction remains strongly significant. This suggests that the NREGA effect seems not to be picking up a moderation in the Monsoon shock and conflict relationship due to the presence of a large mineral resource sector.

A.1.3 Integrated Action Plan

A second important policy aimed to tackle the maoist conflict is the Integrated Action Plan (henceforth, IAP). The plan was presented in 2010 and provides special funding for for districts that are considered to be severely affected by left-wing extremism. Originally it was designed for 33 districts, but since then, it expanded to provide additional funding for 82 districts. The money is to be spend on projects such as roads and other public infrastructure to improve rural livelihoods; some projects are specifically aimed to improving the way NREGA is made accessible in these districts: some IAP funding may be used to complement NREGA projects. Another margin through which the IAP may have a distinct level effect on conflict is provided as money may be used to reinforce police stations to expand the states' presence in rural areas.

Investment in infrastructure funded by the IAP could moderate the rainfall dependence of income and thus, on conflict. I don't think that the IAP would have the effects described in this paper, as its implementation would have to correlate meaningfully with lagged Monsoon rainfall. Since the grants are block grants, this is unlikely to be the case. Nevertheless I study this and the results are presented in Table 6. In any case, there are three simple things I can do to rule out effects of the IAP driving my results. Firstly, I can drop the 33 districts which received the scheme from 2010 onwards.[2] The results from this is presented in columns (2) and (5). The interaction term becomes smaller and size and statistical significance, especially for the conflict intensity regressions. This is not implausible as the districts that receive the IAP are ones with most variation in conflict. In second exercise, I can restrict the analysis to the period from 2000 - 2010. Again, the estimated coefficient on the post NREGA period become weaker, but the core result is still there. In the last exercise I study

[2]The districts translate into 30 districts according to the 2001 Indian census district definitions, they are: Aurangabad (Bihar), Arwal, Balaghat, Bastar, Bokaro, Chatra, Dantewada, Deogarh, Gadchiroli, Gajapati, Gaya, Garhwa, Gondiya, Gumla, Hazaribagh, Jamui, Jehanabad, Khammam, Lohardaga, Midnapore, Nabarangpur, Palamu, Pashchim singhbhum, Purba singhbhum, Rajnandgaon, Rayagada, Rohtas, Sambalpur, Sonbhadra, Surguja, Malkangiri.

IAP fund expenditures, which measures utilisation of the disbursal amounts. Column (1) indicates that IAP expenditures are not correlated with lagged Monsoon rainfall. Column (3) and (6) study the effect of IAP expenditures on conflict. There appears to be a positive relationship between the two. The estimated coefficient on the NREGA interaction remains the same, thus rendering the core result robust.

A.1.4 Operation Green Hunt

A major military operation to tackle maoist violence has been underway since late November 2009. The operation involves the deployment of Central Armed Reserve Police force to aide state governments tackle maoist threat. If deployment of troops is correlated with lagged Monsoon season rainfall, this could explain some of the observed patterns. It is not clear in which direction the effect should be. If military deployment was correlated with lagged rainfall, increased military deployment following an adverse shock could either lead to a conflict escalation or a reduction in conflict. Unfortunately, data on military deployment is not available. As with the integrated action plan, I can limit the analysis to the period before 2010 or by removing a set of districts that likely, were the primary target for a military operation. As these districts are most likely to coincide with districts receiving IAP spending, the results from the previous section alleviate this concern to some extent.

B Data Appendix

B.1 Conflict data

Empirical research on the economics of conflict almost always suffer from severe data limitations. This lies in the nature of the subject of study, that typically places that exhibit conflict are only weakly institutionalised with little official report of violence and little press and media coverage. Blattman and Miguel (2010)'s review cites that the correlation across different civil war datasets ranges from 0.42 to 0.96, which may be the reason why empirical results are often not reproducible using similar identification strategies, but different datasets or variable definitions (e.g. Ciccone (2011)).

There exists no broad conflict dataset that covers India or South East Asia as a whole. This paper documents the process through which in the Indian context 41,347 newspaper reports were transformed into a workable conflict dataset using both machine-learning, semi-automated coding techniques and scalable manual hand-coding methods. The SATP text database represents the most extensive and systematic collection of sources covering conflict in India. It covers an extensive collection of more than 61 distinct English language sources. This vastly improves over the limited set of press agency sources that traditional databases on insurgency violence, such as the Global Terrorism Database (GTD), exploit. Appendix table B1 lists the 61 most frequently mentioned sources; there is a multitude of research papers that have separately hand coded subsets of the primary SATP newspaper clippings, covering various Indian states or various time-periods, see Table B2 for a non-exhaustive tabulation. This results in a coverage of almost ten times as many conflict events.

B.2 Raw sources in the SATP corpus

Table B1: Tabulation of most frequently cited newspaper sources

Rank	News Source	Frequency
1	The Hindu	3862
2	The Times of India	3046
3	Daily Excelsior	3016
4	Times Of India	2663
5	The Shillong Times	2153
6	Nagaland Post	1847
7	Assam Tribune	1752
8	The Telegraph on	1751
9	The Sangai Express	1380
10	PTI	1224
11	Sentinel	1138
12	Imphal Free Press	1131
13	Sangai Express	1092
14	The Sentinel	1071
15	Kangla Online	818
16	Indian Express	728
17	Hindustan Times	706
18	ZEE News	659
19	IANS	650
20	IBN Live	484
21	New Indian Express	453
22	The Business Standard	346
23	The Pioneer	331
24	TripuraInfo	299
25	Kashmir Times	242
26	The Economic Times	215
27	Deccan Chronicle	177
28	ANI	162
29	Sify com	156
30	NDTV	133

This section sketches the semi-automated process through which the daily newspaper clippings are transformed. A typical sample may look as follows:

> Two unidentified terrorists massacred six members of a family and left a seventh injured at Mangnar Top, Poonch district, on December 31, 2001. Local residents refused to cremate the bodies of the slain victims, insisting that a Union Minister should visit the area and take notice of the increasing terrorist violence there.

The semi-automated routine defines a conflict event as a tuple, $E = \{L, T, V, S, O\}$ defined by a location L, a date or time of the event T, a verb V that indicates the type of violent act, and the verb's associated subject S, the perpetrator of the act and the object O that was subjected to the act V. The semi-automated routine tries to fill all these elements of the tuple for each sentence using common machine-learning algorithms implemented in natural language processing packages.

I work with the following set of Trained Natural Language Processing Algorithms:

1. Sentence Detection to break up individual sentences.

2. Semantic Role Labelling (SRL) to tag the grammatical structure of words in relation to one another.

3. Named Entity Recognition (NER) to identify names (places, institutions, names) lives off spelling, preposition and gazetteer. Complemented with dictionary of 1,978 spelling variations.

4. Part of Speech Tagging (POS) to tag role of words (subject, verb, object)

These are together implemented in SENNA (Collobert et al., 2011), available as open-source in C. The sample output for the above sentence would look like:

The refinements include a set of processing steps and can be thought of as a form of supervised machine learning. In the first step, I refine the set of verbs that are indicative of a conflict event. This approach is naturally much more powerful: most hand coding routines remove text in which the word "to kill" does not appear. The set of verbs included is: abduct, ambush, arrest, assassin, assault, attack, attempt, beat up, beaten, beaten up, blast, blew up, blow up, blowing up, blown up, bomb, bombard, boycott, brand, burn, burnt, burnt down, bust, carried out, claim, clash, comb, damag, defus, demolish, desert, detain, deton, encount, ensu, erupt, escap, execut, explod, extort, fight, fire, fled, flee, gunned down, hijack, hit, hurl, hurt, imprison, improvis,

Two			B-A0	B-A0	
unidentified			I-A0	I-A0	
terrorists			E-A0	E-A0	
massacred		massacred	S-V		
six			B-A1		
members			I-A1		
of			I-A1		
a			I-A1		
family			E-A1		
and					
left		left		S-V	
a				B-A1	B-A1
seventh				I-A1	E-A1
injured		injured		I-A1	S-V
at				I-A1	B-AM-LOC
Mangnar	B-LOC			I-A1	I-AM-LOC
Top	E-LOC			I-A1	I-AM-LOC
,				I-A1	I-AM-LOC
Poonch	S-LOC			I-A1	I-AM-LOC
district				I-A1	E-AM-LOC
,				I-A1	
on				I-A1	
December				I-A1	
31				I-A1	
,				I-A1	
2001				E-A1	

infiltr, injur, intimid, kidnap, kill, laid, laid down, launch, lob, loot, lynch, massacr, murder, neutral, neutralis, propag, protest, raid, rape, recov, retali, rob, seiz, set, set ablaz, shell, shot, shot down, shut down, slit, smash, stab, storm, strike, struck, struggl, succumb, suffer, surrend, sustain, threaten, tipped off, torch, tortur, trap, trigger, went off, wound.

This approach maps to the event coding ontology used for global data bases such as the Global Database of Events, Language, and Tone (GDELT).

The second refinement involves exploring the extracted location elements, in order to define a common spatial resolution. In the case of the SATP data for India, the common resolution is the district level. For a significant share of events, finer spatial resolution to the police station level are possible. The population of mentioned locations are mapped and matched against a gazetteer of district names and spelling variations to achieve the geographic resolution at the 2001 census district level.

B.3 Definition of event candidates

A sentence is considered a candidate for an event as long as an object, a verb and a location can be deduced. Information on the date of an event is contained in the meta-information for the newspaper record (the date that it was published). The set of candidates to the set of sentences, for which the location information can be matched to district names and the set of verbs is in the selected subset. In the above text-snippet, only one sentence satisfies the requirement of elements forming an event tuple $E = \{L, T, V, S, O\}$ being present. This yields:

$$\begin{aligned} E_1 &= \{'\text{Mangar Top Poonch}',' \text{December 31 2001}', \\ &\quad '\text{massacre}',' \text{two unidentified terrorists}', \\ &\quad '\text{six members of a family at Mangnar Top, Poonch district}'\} \end{aligned}$$

This is essentially mimicking the process through which humans would code this data manually. Each sentence is constrained to contain only information for at most one event. The individual elements of the tuple E are then transformed by assigning labels to the snippets indicating whether the actor was a terrorist, security force or a civilian and similarly for who subjected to the act V.

In the last step, actors are labeled and conflict casualties are counted. This is done, collecting human judgements from a crowd-sourcing platform and squaring these human judgements with a trained support vector machine learning algorithm. The focus is particularly on the object of a verb. In the above case, we want to label the perpetrator ("two unidentified terrorists") of the act of "massacring" to be "terrorists" and the object of the verb massacre ("six members of a family") to be "civilians."[3]

The dataset has been validated in two ways. First, I cross-validate the resulting data, comparing the machine coding with human coding for a subset of data. The next two sections explain this cross-validation.

[3] Appendix B.1 provides an example with detailed algorithm output.

Table B2: A non-exhaustive tabulation of other papers using or referring to the SATP raw data.

Paper	Time Coverage	Spatial Coverage	Resolution	Primary data source	Method
Shrivastava (2014)	2006-2011	Andhra Pradesh, Bihar, Chattisgarh, Jharkhand, Orissa and West Bengal	District	SATP	Panel regression
Gawande et al. (2017)	2004-2009	Andhra Pradesh, Bihar, Chattisgarh, and Jharkhand	Districts, 2001	local language newspapers	Panel IV regression
Gomes (2012)	1979-2009	16 main states	District	Union across GTD, WITS, SATP	Cross-sectional regression
Hoelscher et al. (2012)	2004-2010	Andhra Pradesh, Bihar, Chattisgarh, Jharkhand, Orissa and West Bengal	District	Intersection across GTD, WITS, SATP	Cross-sectional regression
Khanna and Zimmermann (2017)	2005-2007	Andhra Pradesh, Bihar, Chattisgarh, Jharkhand, Orissa and West Bengal	District, 2001	SATP state time-lines	RDD
Morgan and Reiter (2013)	2001-2011	whole country	District, 2003	GDELT / SATP	Panel regression
Vanden Eynde (2016)	2004-2010	Andhra Pradesh, Bihar, Chhattisgharh, Jharkhand, Karnataka, Maharashtra, Orissa and West Bengal	combined districts using 2001 basis	SATP state time-lines	Panel reduced form regression
Buhaug and Wischnath (2014)	1980-2011	whole country	State	ISPS, SATP, Uppsala	Panel regression

B.4 Comparison with human coding

In order to assess the accuracy of the machine generated conflict dataset, the data for the year 2005, roughly the middle of the sample period, has been hand coded at the sentence level. The human coding can improve along three margins:

1. detect multiple events defined in a sentence
2. assign and infer locations, e.g. provided by names of towns that are not part of a district name.
3. reduce concerns of double counting.

The machine will not be able to detect if multiple events across different districts are mentioned in a single sentence. Here, the machine would, at most, identify a single event for one of the districts mentioned. Similarly, in few instances, the location information is not provided at the district level, but, for example, at the level of town names. In this case, unless the name of the town is part of the district name, the automatic matching to district names, would fail, leaving the location field empty. Lastly, there are concerns about double counting for conflict events that are repeatedly reported on. An example is a terrorist attack which is reported in one week, and a later news report, that adds further information.

Despite these three potential concerns, the machine coding performs extremely well. I assess the fit of the datasets, by simply computing conditional or unconditional correlation coefficients of the number of conflict events at the district level for the whole year (2005), per quarter and at the monthly level.

Table B3: Conditional and Unconditional Correlations Between Human- and Machine Coding

	Time Resolution		
	Year	Quarter	Month
Unconditional Correlation Coefficient	0.987	0.964	0.932
Conditional Correlation Coefficient	NA	0.881	0.826

Human coding does not consistently outperform machine coding. The unconditional correlation coefficients are above 0.95, while the conditional correlation, after removing location and time effects, is consistently above 0.8 (see Table B3).

Visually, the differences are plotted in Figure B1, which presents the conditional correlation, after removing district and month fixed effects, between human and machine coding. The 45° degree line which would correspond to an exact fit. Human coding outperforms machine coding, in cases where information needs to be inferred from the context, which especially applies to coding of location information. However, the overall fit is very good.

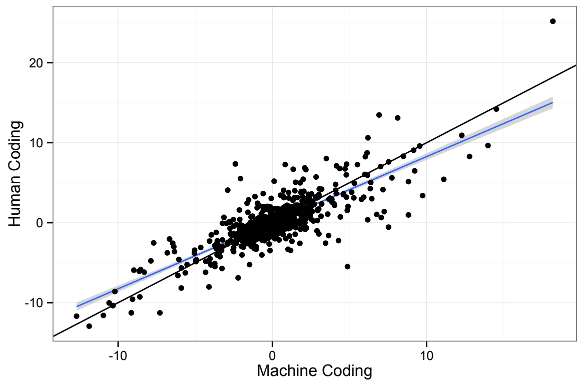


Figure B1: Comparison of Human Coding to Machine Coding for 2005 of conflict event count data at the district level and by monthly resolution. The blue line indicates the linear regression fit, while the black line indicates the 45 degree line.

B.5 Comparison of results with Global Terrorism Database

This section highlights that the results obtained in my paper can not be replicated when studying the conflict for India contained in the Global Terrorism Database (GTD) collected by National Consortium for the Study of Terrorism and Responses to Terrorism at the University of Maryland. This database has been used in more than 30 journal publications and thus, serves as an interesting testing ground. Unfortunately, the GTD database does not come at a district level spatial resolution. However, it provides the nearest big town to where the incident occurred. In order to be able to compare the datasets, I geo-code the locations of the nearest towns to obtain a similar district level count variable of the number of conflict events. I then estimate the main specifications using the number of terrorist incidences in the global terrorism database as a left-hand side. The results are presented in Table B4.

Columns (1)-(3) study the dataset used in this paper, while columns (4)-(6) use the GTD database. In column (4) it becomes obvious that in the GTD data, there appears to be no statistically significant correlation between rainfall and conflict, while there is a strong documented in the dataset used in this paper, indicated in column (1). The geographic coverage of the GTD dataset is a lot more limited before the introduction of NREGA, with only 57 districts reported as having violent incidences before NREGA was introduced while there are almost three times as many districts reported in the

Table B4: NREGA Effect in the GTD and this dataset

	This Dataset			Global Terrorism Database		
	(1) Pre NREGA	(2) Dynamic	(3) Level	(4) Pre NREGA	(5) Dynamic	(6) Levels
Monsoon	-0.866***	-1.330***	-0.680***	-0.985	-1.338*	-1.062**
	(0.270)	(0.306)	(0.261)	(0.684)	(0.764)	(0.462)
NREGA x Monsoon		1.098***			0.359	
		(0.388)			(0.676)	
NREGA			-0.540***			-1.098
			(0.166)			(1.264)
Observations	2841	8868	10199	851	5268	5268
Number of Districts	148	217	217	57	186	186

Notes: All regressions are estimated using a pseudo-maximum likelihood estimator, whose moment conditions coincide with a Poisson model. Regressions in columns (1)-(2) and (4)-(5) include region-phase-time effects as well as district fixed effects, while results for columns (3) and (6) come from a regression with time- and district fixed effects. The dependent variable is the number of incidences per district and quarter. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

other datasets. The moderating effect of NREGA is seen only in column (2), but not in column (5), albeit the coefficient is positive.

As the number of districts covered in the GTD database seems to increase significantly when expanding the analysis to the whole time-period in column (5) it becomes instructive to study how the correlation between these two datasets has evolved over time. I regress the two datasets onto one another, allowing for there to be a separate coefficient for each year:

$$GTD_{dt} = \delta_d + b_{rt} + \sum_{t=2000}^{2010} \gamma_t A_{dt} + \epsilon_{dt}$$

The estimated coefficients γ_t are plotted out in Figure B2.

The specification, by using district- and region by time effects takes out any fixed-conflict region and time varying reporting differences, while the district fixed effects remove any time-invariant district specific reporting biases. The coefficients paint a very stark picture: the datasets do not compare well at all before 2007. The good news is that the coefficients are consistently positive, suggesting that the overall correlation is positive. However, the point estimates are very small and only sometimes statistically significantly different from zero. This suggests that int he earlier years it is extremely unlikely for an incident captured in one dataset to appear in the other. In more recent years, the data become increasingly similar.

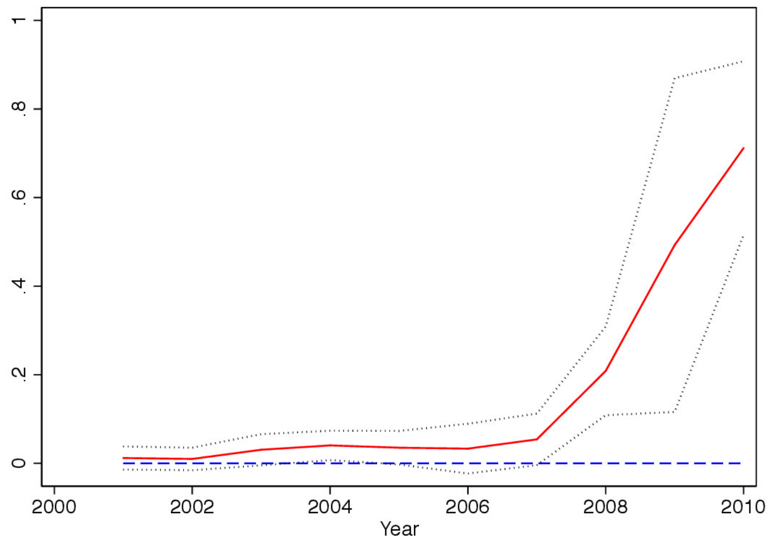


Figure B2: Relationship between this data and GTD Data over Time

Why have the two datasets converged? It appears that the underlying data source in the GTD database has evolved significantly over time. Since 2008, the SATP reports feed into the GTD database, while before that the GTD database was mainly fed by newswire services. By 2010, more than 53% of the incidences in the GTD database were directly referenced with a report from the SATP newspaper clippings dataset. This is clearly, a lower bound since for many reports in the GTD dataset one can manually find references in the SATP dataset, but not necessarily vice versa.

While the level of violence reported in the GTD database seems to be significantly lower for early years, it is important for the identification whether this mismatch in reporting is correlated with rainfall realizations.

In order to explore this, I measure the differences and the absolute value of the differences between the two datasets and run the three specifications from above again.

The results are presented in table B5. The coefficients suggest that a positive rainfall realisation in the preceding month is significantly correlated with a lower reporting difference, i.e. implying that the mismatch between the data used in this paper and the GTD dataset is smaller. This highlights that reporting is likely to be endogenous to past weather and thus, past income realizations. While this is something that can fundamentally, not be checked, I believe that this is more likely to be a problem for the GTD database. The introduction of NREGA appears to have further reduced the mismatch between the two datasets.

If we take this and the previous results together, this suggests that there is some

Table B5: Evolution of Reporting Differences between GTD and this datasets

	Reporting Difference			Absolute Value of Reporting Difference		
	(1) Pre NREGA	(2) Dynamic	(3) Level	(4) Pre NREGA	(5) Dynamic	(6) Levels
Monsoon	-0.078**	-0.090**		-0.107***	-0.136***	
	(0.032)	(0.036)		(0.030)	(0.034)	
NREGA x Monsoon		0.051			0.060	
		(0.042)			(0.043)	
NREGA		-0.398	-0.048		-0.503*	-0.094*
		(0.269)	(0.055)		(0.278)	(0.050)
Observations	12657	25521	27693	12657	25521	27693
Number of Districts	543	543	543	543	543	543

Notes: All regressions are simple linear regressions with time- and district fixed effects. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

systematic differences to the GTD dataset which correlates with rainfall in a systematic way and the introduction of NREGA may have lead to a moderation of this reporting difference. Since the two datasets appear to be converging over time and the coverage of the GTD dataset actually expanding, it seems reasonable to conclude that the SATP data source on which my dataset is a more consistent way to measure conflict.

B.6 Agricultural output, harvest prices and District Domestic Product

I construct a measure of district level agricultural output. This is coming from crop production statistics collected at the district level, by the Directorate of Economics and Statistics in the Ministry of Agriculture.[4] Its the highest quality district level crop production database that exists for India at the district level. District offices collect the data and then forward the data to the Directorate of Economics and Statistics. Typically, there are reporting lags of one or two years. The data is of high quality, but does not reflect total agricultural production. Production from small scale farms or home production is not included.

For every district, I compute a measure of agricultural output value. I only consider crops that have been consistently planted on at least 1000 acres for the whole period that the state reports agricultural production. This leaves the following crops: bajra, barley, castor-seed, chilly, cotton, gram, groundnut, jowar, jute, linseed, maize, mesta, potato, ragi, rapeseed, rice, sesamum, sugarcane, tobacco, tumeric, tur-arhar and wheat. These capture India's most important staple crops as well as cash crops. Underrepresented is production of fruits or other horticulture products.

[4] This data is available on http://apy.dacnet.nic.in/cps.aspx, accessed 14.12.2013.

For each of these crops, I obtained state-level farm harvest prices to compute a district level measure of the agricultural output value. Unfortunately, district level harvest prices were not available throughout or only for a limited number of crops that did not match well with the actual planted crops. For that reason, I stuck with the state-level prices. The resulting dataset is an unbalanced panel going from 1999 to 2011.

Total output produced is significantly lower than total district level agricultural sector output value. For the quantification exercise on the insurance value, I scale up the district level agricultural output value to match the district domestic product for the year 2000. The district domestic product is an estimate of local area incomes that has been produced for the period 1998-2005, but is not available for more recent years.[5] It relies on a large set of input statistics, including the Annual Survey of Industry, the National Sample Survey and Crop Production Surveys. The district domestic product construction is discussed in detail in Katyal et al. (2001). I obtain a baseline measure of the agricultural output per capita from the district domestic product. This measure will be unambiguously larger than the computed agricultural output value derived from the crop production statistics, as I only include crops that have been consistently reported for the time period that a state reports data to the Directorate of Economics and Statistics. I compute for each district a scaling factor ω_d that measures the share of the agricultural output value per capita that is captured in the agricultural district domestic product. I then simply scale up the agricultural output value per capita by this scaling factor. This preserves the variation but likely gets the agricultural output value closer to the true. This scaled agricultural output value per capita will be used for the quantification exercise to evaluate how much insurance NREGA provides.

B.7 Agricultural Wages in India

I use wage data from the annual publication Agricultural Wages in India, which has been published from 1958 up to 2010. This data provides up to monthly frequency, village level wage data across Indian districts. This is the raw data that has been used to construct the World Bank Agriculture and Climate dataset on India that has been very widely used, most recently, by Kaur (2018). For the period 2005-2010, versions of the data can be obtained in electronic form[6]. Data for previous years does not exist in digital form, but needs to be obtained as hard copies from India and then digitized. I

[5] The data is available from http://planningcommission.nic.in/plans/stateplan/index.php?state=ssphdbody.htm, accessed on 21.06.2014.

[6] This is available from http://eands.dacnet.nic.in/AWIS.htm, accessed 12.10.2012.

obtained hard copies of the reports and digitized the data from 1998 onwards.

The raw data gives monthly wages for male, female and children, broken into skilled- and unskilled agricultural labour and different types of labour. The types of skilled labour are blacksmith, carpenter and cobbler, while unskilled labour combines ploughman, reaper/harvester, sower, weeder, other agricultural labour. In some states, these separate unskilled labour categories are not reported[7], but rather, a category "Field Labour Wages" is reported. I focus on the male wage series for the occupations "ploughman" or "field laborer"

In some districts these wages are reported throughout the year, while in others the wages are reported only in the parts of the year, when particular activities are actually carried out (i.e. sowing wages in the early monsoon season of May, June and July), while harvesting wages are reported in the fall of a given year.

After digitizing and entering the raw data, I proceed to construct an annual level agricultural field-labour wage as my main dependent variable as simple average. This results in an unbalanced district level panel stretching from 1998 to 2010.

B.8 TRMM rainfall data

This paper is the first one in economics to use data from the Tropical Rainfall Measuring Mission (TRMM) satellite, which was jointly operated by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace and Exploration Agency (JAXA). The satellite carried a set of five instruments to construct gridded rainfall rates at very high spatial and temporal resolution. It recorded data from 1998 to 2014, when the satellite started to de-orbit. The dataset is considered to be the highest quality remote sensed rainfall dataset, with global coverage, currently available (see Li et al., 2012).

The TRMM Multi-Satellite Precipitation Analysis provides daily rainfall from 1998 to 2014 at a fine spatial resolution of 0.25 by 0.25 degree grid-cell size. The data from the various instruments aboard the satellite are cleaned and calibrated using additional data from the accumulated Climate Assessment and Monitoring System (CAMS). The output of the algorithm are 3-hourly rainfall rates for that time-period. This is then scaled up to obtain monthly mean precipitation rates, which in turn are transformed into overall monthly rainfall.

Remotely sensed weather data is an important source of data, in particular, for less developed countries, where observational data is scarce. This is particularly relevant

[7] The states for which this is the case are Andhra Pradesh, Karnataka and Maharashtra.

in the case of India, where observational weather may vary in systematic ways. There are three main drawbacks. First, most observations come from rain gauges, where measurements are taken once a day. Climatologist are concerned about rain gauges in particular in tropical- or subtropical areas, since most rainfall is convective. Such convective rainfalls are highly local, generating intermittent and scattered rainfall, which may not be picked up using rain gauges, if the network is not spatially fine enough. The TRMM satellite orbits the earth every 90 minutes, thus providing multiple observations each day. An alternative is to consider data from weather radars. Rainfall radar may provide estimates for rainfall in a radius of 200 km around the station, however it is unreliable for distances in excess of 200 km. In the Indian case, rainfall radar data is not made available and would be problematic, since most reporting radar stations are clustered along the coast. The third general concern regarding observational weather data is the fact that reporting may be endogenous e.g. to violence or other variables that are correlated with the dynamics of violence. This has been highlighted recently by Smith et al. (2011), who show that Somalian piracy has generated a "black hole" in the Indian ocean, where observational weather data from merchant vessels is not available anymore, as vessels take routes avoiding piracy infested areas.[8]

I prefer the TRMM data as it is less subject to systematic measurement error, as the underlying data source is consistent over time. This is not the case with rain gauge based data, such as the GPCC as used by Miguel et al. (2004), Ferrara and Harari (2018) and Kudamatsu et al. (2014) and many others. In the case of India, the number of reporting weather stations for the GPCC data set varies from year to year. In 2001 there were a total of 1197 stations that reported at least some data, while in year 2008 that number dropped to 978. On average, 15.7 % of the district-year observations have some rainfall station reporting data. This pattern varies systematically with violence as is shown in table B6. The table presents results from the same specification as in the main part of the paper, including region-by NREGA phase time effects and district fixed effects. The dependent variable is an indicator whether any station reported data for that district and year. The regressor is either an indicator whether a district experienced any violent incident in the last year (column (1)) or the number of incidents in column (2).

The coefficient on the violence indicator is insignificant, with a p-value of 18.5%. The coefficient on the number of attacks is significant at 5%, indicating that one addi-

[8]Another example is the case of Vanden Eynde (2016), who had to merge several districts together in order to obtain consistent rainfall estimates, since many stations simply fail to report rainfall estimates. Most of these stations are located in places with conflict or in newly created districts or states.

tional attack per year decreases the probability of a rain gauge station reporting data in the subsequent year by 1.3%, when evaluating it against the mean of the dependent variable. Despite this general concern, my results are robust to using either the GPCC data (Schneider et al. (2011)) or the Indian Meterological Department data used in Vanden Eynde (2016).

As both the GPCC and TRMM data have been processed using climatology algorithms, a general concern is "error propagation" (see Leung et al., 2005; Burnicki et al., 2007). As the raw data is transformed in the analytical process, the mathematical and numerical transformations may propagate small simple measurement errors. This could generate spurious correlations that could affect the results. I remove this systematic and non-systematic measurement error by instrumenting one dataset for the other.

Table B6: Weather Station Reporting in GPCC Varies with Violence

	(1)	(2)
Any Violence	-0.013	
	(0.009)	
Attacks		-0.002**
		(0.001)
Mean of DV	.157	.157
Observations	5440	5440
Number of Districts	544	544

Notes: All regressions are simple linear regressions with time- and district fixed effects. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.9 NREGA data sources and roll out

The data for the roll-out of NREGA come from the Ministry of Rural Development, which is responsible for administering the scheme. The sequence of roll-out was highly endogenous to a set of district level characteristics, such as the share of scheduled caste, scheduled tribe population, baseline agricultural productivity, literacy and existing levels of conflict. This becomes obvious when considering Figure 1. This picture highlights that a lot of districts in the east of India received NREGA in the first round. A lot of these districts did suffer from maoist violence. As discussed in the main body, I do not require exogeneity of treatment to levels of violence for my empirical design. There are two main sources for data on NREGA take-up. These are the district-level monthly-progress reports (MPR) and data coming from the Management Information System (MIS). The latter is a completely non-paper based system that has only become mandatory to use in the financial year but was still not fully operational until 2010-2011.

There are a lot of issues regarding the reliability of either datasets, as there is quite some mismatch between the two datasets, especially in the earlier years when the MIS was introduced.[9] This may be due to partial compliance in the MIS after it had been introduced, but could be also because the MPR system is more subject to manipulation. It is difficult to asses the underlying divergence in the two databases.

The MPR data is available continually from 2006 to the financial year 2010-2011, from which point onwards I rely on data from the MIS.[10] The format of the reports has changed considerably, with the major break occurring in 2011. This is partly due to the evolving nature of NREGA. Ministry of Rural Development (2009) details that several programs by the Ministry of Water Resources are to be joined with the NREGA by 2011. An important part of this program are rural sanitation projects that are funded by the Ministry of Water Resources for a set of targeted districts. This implies that there are district-specific breaks in the NREGA data. In the empirical specifications which combine data from before and after 2011, I flexibly control for these breaks by allowing the district fixed effects to be different before and after 2011.

In addition, there are several variable name changes, which may not necessarily reflect identical concepts. For example in the MPR data, "No of Families Completed 100 days" while in the MIS data the variable is labeled "Total households reached 100 day limit". Similarly, the MPR data contains a variable "Total Employment Demanded", which, studying levels, seems to reported by number of households. In the MIS data, this variable is labeled more clearly as "Total households demanded work". Since the data is collected at monthly frequency, it is not clear whether the numbers refer to the total number of unique households who demanded employment within a month, or across months.

I focus on the set of variables that is most cleanly and most consistently reported: expenditure per capita, number of person days worked per capita as well as the total number of households demanding employment.

Despite having access to NREGA for many months in a financial year, I only study the reported metrics at the end of each financial year (that is March of each calendar

[9]See for example mismatch between MIS data and National Sample Survey returns data highlighted by http://www.indiatogether.org/2013/jun/gov-nregs.htm,accessed on 12.06.2013.

[10]Thanks to Clement Imbert for sharing NREGA MPR data for the earliest years.

year). This becomes necessary as there are significant reporting delays which induce large jumps in the cumulative month on month measures which are less likely driven by participation, but more likely due to reporting issues. This measurement error was most pronounced in the period for which only the paper based MPR based data is available. The take-up pattern is clear in that the bulk of the NREGA participation is in the lean season, directly following harvest. See for example for Andhra Pradesh, Figure 3.

C Additional References for Appendix

Aggarwal, S. (2014). Do Rural Roads Create Pathways Out of Poverty? Evidence from India. *mimeo*.

Blattman, C. and E. Miguel (2010). Civil War. *Journal of Economic Literature 48*(1), 3–57.

Buhaug, H. and G. Wischnath (2014). Rice or riots: On food production and conflict severity across India. *in Press, Political Geography*.

Burgess, R. and D. Donaldson (2010). Can Openness Mitigate the Effects of Weather Shocks? Evidence from India's Famine Era. *American Economic Review: Papers & Proceedings 100*(2), 449–453.

Burnicki, A. C., D. G. Brown, and P. Goovaerts (2007). Simulating error propagation in land-cover change analysis: The implications of temporal dependence. *Computers, Environment and Urban Systems 31*(3), 282–302.

Ciccone, A. (2011). Economic Shocks and Civil Conflict: A Comment. *American Economic Journal: Applied Economics 3*(4), 215–227.

Collobert, R., J. Weston, and L. Bottou (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research 12*, 2461–2505.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Ferrara, E. L. and M. Harari (2018). Conflict, Climate, and Cells: A Disaggregated Analysis. *The Review of Economics and Statistics 100*(4), 594–608.

Gawande, K., D. Kapur, and S. Satyanath (2017). Renewable Natural Resource Shocks and Conflict Intensity. *Journal of Conflict Resolution 61*(1), 140–172.

Gomes, J. (2012). The Political Economy of the Maoist Conflict in India: An Empirical Analysis. *mimeo*.

Hoelscher, K., J. Miklian, and K. C. Vadlamannati (2012). Hearts and mines: A district-level analysis of the Maoist conflict in India. *International Area Studies Review 15*(2), 141–160.

Katyal, R., M. Sardana, and D. Satyanarayana (2001). *Estimates of District Domestic Product*. New Delhi: Socio-Economic Research Centre.

Kaur, S. (2018). Nominal wage rigidity in village labor markets. *forthcoming, American Economic Review*.

Khanna, G. and L. Zimmermann (2017). Guns and butter? Fighting violence with the promise of development. *Journal of Development Economics 124*, 120–141.

Kudamatsu, M., T. Persson, and D. Strömberg (2014). Weather and Infant Mortality in Africa. *mimeo*.

Leung, Y., Y. Ge, J. Ma, and J. Wang (2005). Measurement Errors and their Propagation in the Registration of Remote Sensing Images. In *Developments in Spatial Data Handling*, pp. 285–297. Springer Berlin Heidelberg.

Li, X.-H., Q. Zhang, and C.-Y. Xu (2012). Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. *Journal of Hydrology 426-427*, 28–38.

Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy 112*(4), 725–753.

Ministry of Rural Development (2009). Joint Convergence Guideline. Technical report, Ministry of Rural Development.

Morgan, R. and D. Reiter (2013). How War Makes the State: Insurgency, External Threat, and Road Construction in India. *mimeo*.

Schneider, U., A. Becker, P. Finger, B. Meyer-Christoffer, Anja; Rudolf, and M. Ziese (2011). GPCC Full Data Reanalysis Version 6.0 at 0.5 Degree: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data.

Shrivastava, A. (2014). Civil conflict with rising wages and increasing state capacity: The Maoist insurgency in India. *mimeo*.

Smith, S. R., M. A. Bourassa, and M. Long (2011). Pirate attacks affect Indian Ocean climate research. *Eos, Transactions American Geophysical Union 92*(27), 225.

Vanden Eynde, O. (2016). Targets of Violence: Evidence from India's Naxalite Conflict. *The Economic Journal* (November), 1–50.